DOCUMENT RESUME

ED 176 779 IR 007 736

AUTHOR	Eva[illegible], Glyn T.; And Others
TITLE	Development of a Responsive Library Acquisitions Formula. Final Report.
INSTITUTION	State Univ. of New York, Albany. Central Administration.
SPONS AGENCY	Bureau of Libraries and Learning Resources (DHEW/OE), Washington, D.C.
BUREAU NO	475AH70078
PUB DATE	Dec 78
GRANT	G007702305
NOTE	107p.; Examples may not reproduce
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Academic Libraries; Data; Data Processing; Library Acquisition; *Management Information Systems; Mathematical Formulas; Taxonomy
IDENTIFIERS	*Collection Development; Higher Education General Information Survey

ABSTRACT

The major accomplishments in the development of a management system for academic library collection development were (1) the development of translation tables which express HEGIS (Higher Education General Information Survey) taxonomy terms as sets of LC (Library of Congress) class numbers; (2) the use of these tables to compare library acquisition, institutional administrative data, and book publishing data, and to treat them as components of a library management system; (3) the development of computer programs to build and maintain the tables, isolate and analyze the data, and prepare a series of reports which together comprise the library management information system; and (4) the establishment of a process by which libraries and institutions can analyze their data. The attempt to develop quantitative components which would be included in a proposed acquisition formula failed, although the structure of the formula appears to be validated. Appendices present the modified HEGIS taxonomy, a list of subject specialists who compiled tables, instructions for the assignment of LC classification numbers, instructions for proofreading LC/HEGIS tables, an example of an LC/HEGIS conversion table, and the proposed SUNY (State of New York) formula for current acquisitions. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED176779

Final Report

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Project No. 475AH70078

Grant No. G007702305

DEVELOPMENT OF A RESPONSIVE LIBRARY ACQUISITIONS FORMULA.

FINAL REPORT

BEST COPY AVAILABLE

Glyn T. Evans, Project Director
Mary H. Beilby, Research Team Coordinator
Roger D. Gifford, Programmer/Analyst

State University of New York Central Administration
Office of Library Services
State University Plaza
Albany, New York 12246
(518) 474-1430

The activity which is the subject of this report was supported in whole or in part by the U.S. Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U. S. Office of Education, and no official endorsement by the U. S. Office of Education should be inferred.

IR007736

A Report on a Grant Awarded By the

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
Office of Libraries and Learning Resources

ABSTRACT

This report describes developmental work on a management information system (MIS) for academic library collection development. The system will contain data required by library collection development personnel, library managers and academic administrators to develop library collections. Data requirements center around what the library is collecting, what it should collect to meet the needs of library users and what funding is necessary to buy this material. The data used in the information system are primarily generated as by-products of other operations such as library on-line cataloging and acquisitions, listings of publishing output, and academic program information produced for general university management.

The data are aggregated into instructional subjects which allow library and publishing data to be related with academic data. Since courses are classified in the HEGIS taxonomy, and academic library books are commonly classified by the Library of Congress Classification, a mapping of these two classifications to selected instructional subjects allows a direct comparison of library collections and course offerings. A crosswalk or translation table between HEGIS Taxonomy and LC Classification has been constructed, in which some 12,000 LC ranges describe some 500 HEGIS terms.

The management information system is incomplete. Its current state of development allows the three groups of data to be processed in a variety of ways. For example, data from two or more files may be integrated in one report, and files may be manipulated using the Statistical Package for the Social Sciences. Libraries may acquire either printed reports or processed data in machine readable format. A few descriptive reports are presented to demonstrate the capability and viability of the MIS. Initial attempts at constructing decision models are shown.

The attempt to establish generalized values in support of a proposed library acquisition formula was not successful, although the structure of the formula appears to be valid. Instead, it is now proposed that through use of the structure combined with this library management information system it is possible for each individual academic institution precisely to define and budget its own acquisition needs.

Directions for further research are outlined.

CONTENTS

		PAGE
1.	Background	1
2.	The Results of the Project	3
3.	Data	10
4.	LC/HEGIS Translation	18
5.	Project from a data processing perspective	27
6.	Reports generated	38
7.	Summary and Conclusions	68
8.	References	69
9.	Acknowledgements	71

Appendixes

1.	CASA/HEGIS Taxonomy	73
2.	List of Subject Specialists Who Compiled Tables	79
3.	Instructions for the assignment of LC Classification Numbers	81
4.	Instructions for proofreading LC/HEGIS Tables	84
5.	Example of LC/HEGIS conversion table	88
6.	Proposed SUNY Formula for current acquisitions	90

1. BACKGROUND

This project builds upon a previous project: "Collection Development Analysis using OCLC Archival Tapes" funded by the Office of Education, Library Research and Demonstration Program (Project no. 475AH60088, Grant no. G007603346) (Evans, 1977).

A major outcome of that project was a set of computer programs for analyzing OCLC archival tapes and other library data sources. Eleven programs were written for the first project; three extract and manipulate bibliographic information from OCLC-MARC records which is analyzed by the other programs.

The primary consideration of the analysis programs was that they utilize a variety of data, in fixed-length records in order to process OCLC tapes, circulation system files, master files from the publishing trade, keypunched shelflist samples, etc. Furthermore, because the data being processed are described to the programs by a table of field definitions, the programs are not limited to library data analysis. They could also process accounting or budgeting system data, or institutional data (such as course enrollment).

A second major outcome of the 1977 project was a new methodology for analyzing library collection data. Content Analysis, accounts for all titles in a library collection by partitioning the collection into discrete segments. For example, the A.L.A./R.T.S.D. Collection Development Guidelines (1976) break the LC Classification schedule into 482 discrete segments with 92 intermediate and total cumulations.

A second type of analysis, Component Analysis, focuses on a specific "component" of the collection and counts the number of titles in the collection relevant to that component. To the researcher or library manager, a "component" may be a subject, a discipline, the course content, research area, the scope of an academic department or program, etc. Operationally, a component is expressed as one or more single values or ranges of values in a classification scheme. In an LC-classified collection, these would be expressed as LC call numbers, in a Dewey collection by Dewey, class numbers in a medical collection by NLM, etc.

Components may overlap. For example, both the Biology and Chemistry components of a collection might contain Biochemistry classification numbers. These classifications would thus define a Biochemistry component if Biochemistry is a component of interest.

Components may also be arranged in a hierarchy, with the most comprehensive component at the top of the hierarchy and the most specific at the bottom. There is no limit to the number of levels in such a hierarchy.

One problem with overlapping hierarchies of components is that the number of titles in a component will be less than the sum of titles in its subcomponents because one title may be contained in more than one component.

Example:

	Titles
Biological Sciences	100
Zoology	50
Botany	50
Bacteriology	10
	110

No title is ever counted more than once in a component no matter how many of its subcomponents or ranges contain it. This is an integral feature of the programs and is a counting method not available in other statistical programs.

The third type of analysis defined in the previous project, is Overlap Analysis. Overlap Analysis describes the type and amount of title overlap existing among components. One way to do this is to list the titles held by more than one component; Another is to count the number of titles held by one component, the number held by two, and so on. A third method, which shows the percentage of titles each component holds in common with other components. Such displays are useful only if the components being compared are at approximately the same level in the hierarchy, e.g. subjects of approximately the same scope or libraries of comparable size.

Although the programs and types of analysis developed in the previous project were designed for library data, they are applicable to other areas as well, and it is this property that lends itself to the current project. Content, Component, and Overlap Analysis can be used to analyze the courses offered by an academic institution, or the patrons of libraries by their subject or area of interest.

The concept of a hierarchy of components is completely general. It can be used to represent the campus structure of a multi-campus university, the department structure of a campus, the composition of a collection by language, subject or country of publication, or the composition of a campus population by language or country of origin.

One of the most important aspects of this generality is that it allows the study of trends. If one considers the acquisitions of a library for a single day, month, or year as a component (with time as the classification scheme), then the capability to perform Content, Component, and Overlap analyses over time becomes apparent.

2. THE RESULTS OF THE PROJECT.

The title of the project "Development of a Responsive Library Acquisitions Formula" emerged from work which had been undertaken at SUNY Central Administration Office of Library Services on Acquisitions formula development. A formula which moved away from the enrollment-based formulae exemplified and initiated by Clapp and Jordan was proposed in 1976. The SUNY formula turned primarily to the discipline and its record rather than to the campus and its population. The first grant on collection development analysis (Evans) was in progress and the early results obtained by the time this present project proposal was being written encouraged me to believe that it would be possible to link the two actions together, and in fact derive the values for the SUNY formula through the programs being developed for collection analysis.

In the event, the result is that the structure of the proposed formula has, by and large, been validated but that the rather blunderbuss allocations of the original proposal have been discarded.

The result is a system which is at the same time more subtle and more satisfying.

a) Formula Budgeting for Acquisitions

It is reasonable to ask the question whether or not it is possible to develop an acquisition formula. Formulae in higher education have had an interesting history and have been a mixed blessing. Library acquisitions are not the only expenditure and planning activity which have been the subject of formula development. A review of Halstead (1974) 'Statewide Funding for Higher Education' reveals a wide variety of budgetary and planning formulae for space facilities, class size, enrollment, course planning, as well as for libraries. It is generally admitted that formulae are attractive to administrators and budget officers alike. Provided the formula is well defined, rooted in sound data, and understood by all parties, a formula can remove much of the guess work out of both current activity and forward planning. In short, a formula is an attractive political as well as administrative expedient. In a review of formula budgeting in a paper 'Formula Budgeting: requiem or renaissance?' Moss and Gaither (1976) conclude that formulae are indeed valuable, but they need to be extremely sensitive to local conditions and variations over time.

The Clapp/Jordan formula was welcomed and used by many institutions, as was its derivative, the Washington formula. It has fallen into disfavour in part because analysis demonstrated that it was not statistically verifyable (McInnis, 1972) but also, and primarily, because it established the concept of 'minimal adequacy', which budget officers used to contain or cut acquisition budgets. It was recognized that there is political value in establishing a target for absolute size, but what happens when that target is reached? The practical, budgetary answers to that question have severely cut some institutional budgets.

In fairness, it should be pointed out that Clapp/Jordan was preliminary work, and that the writers both cautioned against unthinking adoption and identified many further areas for research, few of which were followed up. We do grasp at straws.

The formula problem was recognized early in SUNY and a number of attempts were made to propose alternatives, most of which, unfortunately,

proposed the same structure with higher numbers, a tactic which at best could only buy time. In 1972-3, however, the writer participated in work performed by a SUNY-wide Faculty Senate Library Committee study on 'Factors Affecting Library Collection Development' (1974) which placed primary importance on the 'discipline'. Projects mounted at SUNY have continued to emphasize and refine that stress.

b) Proposed SUNY Formula: 1976

The proposed formula is attached as Appendix 6 to this report. It is a formula for current acquisitions only, and does not, with one exception, address retrospective acquisition. Distribution of the proposal to campus presidents and head librarians for review resulted in general consensus as to the form and structure of the formula but no specific endorsement of the actual numbers allocated, with one exception, was made, and rightly. The actual numbers proposed were based on reviews of published output as best as could be determined, and on estimations of campus demand based on brief summary campus data, an immediate recognition of the paucity and uncertainty of the data. The allocations were also, frankly, influenced by the size of the (then) current budget allocations. It would, clearly, have been unacceptable to propose allocations which would have reduced the library budgets. Apart from the political ramifications of such a proposal, historical reviews of campus budgets going back to 1967, (When the Bowker Price Indexes were established at 100), revealed severe degradation in the budget allocations compared with the inflationary pattern of price increases, with also, of course, the concomitant affects of reduced numbers of current acquisitions and a violent swing in the internal allocation of resources from monographic to serial materials.

The one exception to the numbers cited above was the case of the University centers, which had at that time expressed support for the Voigt formula. It was found that the Voigt formula could be expressed in the same terms as the proposed structure and hence the Voigt numbers were used for those four libraries.

Given the general acceptance of the structure, the problem became one of establishing allocations. That need, and the development of the collection analysis programs, led to this project.

The fundamental structure of the formula is:-

a) A base collection (used to maintain reference collections and to round out or balance the total collection; also should be used to support external data access, e.g. on-line retrieval)

plus

b) Disciplinary allocations based on:-

a) the academic mission or program of the institution, including research interests

and b) the level at which the disciplines are offered at the institution, i.e. 2 year, baccalaureate, Masters, Ph.D., Professional.

and c) modified by the publishing output for discipline, and also by form (monographic, serial) format (paper, maps, microform, phonodisc, av,) and such other factors that influence budgetary allocations (e.g. number of foreign publications required) as could be identified.

plus

d) Increments on base to accommodate campuses with larger enrollments, where the need for investment in multiple copies of certain titles is obvious and necessary

The philosophy expressed by this structure is that if the campus is offering a program, it is making a commitment, or investment, in resources to support that program regardless of the numbers of students, faculty or researchers who will participate in the program, and that this requirement is identifyable and assessable. Further, it recognizes the need for multiple copies as a function of volume of demand for some materials and some items.

The problems of actual use of the materials are not included in the formula, or this project. They are, however, recognized in a subsequent grant request in which collection use is proposed as the essential feedback mechanism which informs and illuminates the acquisition/selection decision process. This aspect was, of course, recognized before the current project began, but sometimes small bites are best.

e) Results of the project.

It should be said at once that the project has not established values which can be plugged into the structure outlined above, even though it was hoped that that could be done. It has developed a management information system which is, it is believed, a considerable improvement over the standard approach proposed in the formula.

The key to the problem is the wide and subtle variety of the individual campus need. It is now recognized as a matter of historical fact that library acquisition formulae have evolved from general assessments (based on historical analysis and some empirical studies) generally applied to more sensitive proposals which recognize variation and complexity. The development from Clapp/Jordan to Voigt is one example. Further, it is reasonable to assume that this evolution has been accompanied by, and probably made possible because of, a general increase in the availability and accuracy of management data as well as the natural growth of personal understanding.

The range of available management information data has been considerably increased by this project, as is shown in this report, and that is its first major success. The creation of the interface tables between academic, library collection, and publishing data (and one hopes in the future collection use data) has opened new doors to understanding.

The second major success has been the establishment of a battery of computer programs which build on the earlier work but are extended to accommodate a wider variety of data and have, in some cases, been revised

to improve computing efficiency. Further, it is important to recognize that although they are limited to certain data sources, e.g. ARPR, there seem to be no reasons why other data could not be incorporated in the system. The development and acceptance of UNIMARC, for example, and the promise it holds for Universal Bibliographic Control (UBC) is welcomed enthusiastically by the project (LC Infor Bull). The broader the base of data the better the library management information system and the more applicable its results.

It remains for the project to be turned into an ongoing program but SUNY Central is continuing to invest internal resources to assume the establishment of an ongoing program. The work initiated in this grant project must not be allowed to lapse.

The third aspect of the success of the project is the wide variety of reports which can be generated, the questions they answer, and the questions they raise. It is a disappointment that within the operational parameters of the grant, the sheer magnitude of the tasks undertaken and accomplished did not allow extensive analyses of data from many institutions which might have answered many of the urgent questions facing libraries. Regretably, all one can honestly say is that some reports have been run on some data but nowhere nearly enough to draw any operational conclusions. There, again, the continued investment of effort by SUNY is important. It is also fair to note that a good deal of interest has been generated within the profession following reports of 'work in progress' by the project staff, and that it is likely that SUNY will be requested to process data for other institutions.

These analyses, from a wider base, should throw further light on the problems of acquisition patterns and needs. Reports of significant results will be made available to the wider professional community.

d) Assessment

It is reasonable to ask what our understanding of the problem of acquisition formulae is now.

1. The project believes that instead of a formula it is now possible to establish a structured management information system which is institution-specific, and that this is a more important step than the original goal of establishing blanket values to be applied across the board to many institutions.

2. The components of the prepared formula have survived, and been validated as necessary components of an information system. Others, specifically collection use data, have been added.

3. Institutions are immensely complex, and the relationship between mission and performance very subtle. Each, however, has its own life, style, goals and resources. The availability of an information system which can inform the local investment process must be a wiser asset than either across-the-board formulae, or management by comparison (i.e. Institution X has N volumes; we are (or want to be) like Institution X; therefore we need N volumes).

Perhaps the best service that can be offered an institution is the ability to assess its library needs in its own terms based on the particular mixture of disciplinary activity embodied in its own mission.

4. The general applicability of the management information system is tied to the extent to which campuses participate in OCLC or similar systems, and the rigor with which they use the NEGIS system. Since it is a phenomenon that more libraries are turning to OCLC, Bellota, etc., and since NEGIS is a federal requirement, one hopes that the process will become generally available and acceptable. Libraries which do not participate in the system are not, of course, disbarred from taking advantage of this work, but they will have to use manual analytic techniques. The fundamental logic underlying this work is not dependent on the existence of machine-readable data.

5. The process does have limitations which are recognized. Some can be incorporated, other not, and it is important to understand what they are.

   a) Data excluded from ASPR:

      foreign publications
      serials/subscriptions
      physical formats (AV/etc.)

      As data becomes available, they can be incorporated. In fact, one product of the analyses is an assessment of those library acquisitions which do not appear in ASPR — a development which has distinct possibilities, particularly in the media area. Despite the exclusions the project feels that it is working with a very high percentage of the monographic acquisitions made by U.S. libraries.

   b) Research collections.

      The major research libraries may invest only a small portion of their acquisition resources in the areas measured by this process. It is for this reason that the results obtained by Kent et al are deceptive. Much work needs to be done to expand his process into a wider range of academic libraries, in part to derive criteria which exclude research collections.

      The investment made in a research collection is presumably intended as much for posterity as it is for current use. There is no doubt that the existence of a "research" need can be used to blur decision making regarding current acquisition. The project proposes that two analyses need to be mounted simultaneously, one using the processes proposed here, the other reexamination in the institutions own terms of those identified research portions of the collection. The project recognizes that the institution may have to continue high investment in some collections even if there is no apparent current need. For example, the sheer monetary value of the collection may

decline sharply if it is not maintained, where the potential loss exceeds the cost of current maintenance investment. Such cases are, presumably, rare. The identification of such collections is probably much more difficult than the current analysis.

c) Retrospective acquisition.

Retrospective acquisition was deliberately excluded from the formula because since in SCSU's terms only three of the campuses were classified as 'emerging' (a temporary state of very high acquisition budgets to build a library when a campus is newly founded - it usually is considered to have 'emerged' after five years), and they would soon be out of that developmental stage, inclusion was of little utility.

Rather, it was expected that a library, or a campus, would identify a need that demanded retrospective redress. For example, a drastic change in program direction, or the establishment of a new department, curriculum, or research area, would be cases where a campus decision forced remedial action by the library. In a well-ordered world the campus would ask the library to identify the cost of investment, the library would identify titles, and serials needed to support the new direction, and the cost associated with each. The campus would understand and support the new investment, and the retrospective collection would be acquired. We do not live in a well-ordered world, and generally the process outlined above does not occur; but budget examiners are increasingly impatient with vague, blanket requests for retrospective acquisition unaccompanied by real statements of purpose, or specifically-identified - and costed - lists of materials.

Of course, a certain amount of retrospective acquisition does occur all the time from current funds - and so it should. The project does propose, however, that items more than three, or at most five years old (other than replacements) should be considered as separate retrospective acquisition.

The programs developed under this project may well be very useful for the identification of retrospective needs after data have been acquired and processed over a number of years.

6. The quality of the data from all three sources (publishing, library, and institution) needs to be improved. It is not surprising that the results so far obtained were a little unexpected. It is unreasonable to expect three such disparate sources to produce instantly homogeneous reports. The natural feedback into the reporting mechanism which will occur given examination of the

reports should improve the data and the understanding of their meaning.

7. One area deliberately excluded from the formula is 'resource sharing' although it is recognized as a vital and necessary library activity. Preliminary data indicates that library collections are a lot richer in variety that previously anticipated. It maybe that the best way to cooperate in library acquisition is to do exactly what we are doing, and rely on delivery systems and such procedures as the OCLC-ILL system to provide access. Many more data from more campuses need to be processed before this suggestion can be supported, however.

   Clearly, as more campuses analyse their data through this project, a corpus of information will emerge which will illuminate the resource sharing process.

e) Summary

The structure of the proposed SUNY formula for current acqusitions appears to have been validated, but it has been superseded by the development a responsive library management information system which will illuminate and support academic institutional acqusitions decision making.

The processing of many data from many campuses may establish norms of acquisition by different disciplines which may offer guidelines, but this remains to be demonstrated.

In any case, library use data must be incorporated as a feedback mechanism in the library information system.

3. DATA

This chapter describes the data used in the current project, the aggregation methods, limitations and restrictions on the data's utility, and describes data which may be available to other academic institutions.

TABLE I

Data categories	Data sources	Specific data elements
Library collections	Library catalogs, OCLC distribution tapes, BALLOTS records, etc.	MARC data elements: language, date, subject classification
Library acquisitions	Machine-readable acquisition & cataloging records	(same as above)
Library use	Machine-readable circulation records	Classification number, publication date, use date, user identification
User satisfaction	Library failure data	Items located/not located by subject, user identification
Publishing output	National bibliographies, Union serial lists, commercial publishing records, e.g., Bowker, Faxon, etc.	Subject classification, title, price, date
Academic programs	University administrative records, HEW statistical reports	Courses, student enrollment, faculty, instructional methods, costs, subject
Research	Faculty Interest Profile, grants awarded	Subject

The data used in the information system are charted in Table I. The data are derived primarily from library, university academic program and publishing industry operational systems. The primary data are in machine readable format such as the MARC tapes or OCLC distribution tapes, the Book Publishing Record and the SUNY Course and Section Analysis (CASA) files. Other institutions will have enrollment and other institutional data in variant forms.

Additional data, not originally in machine readable format can be incorporated into the system by formatting those elements which are required for a particular analysis. For example, the manual library shelflist could be sampled.

There are a number of advantages of using these existing data sources:

1. There is little cost involved in acquiring the data.

2. Since they represent individual transactions, i.e. circulations, books cataloged, books published and carry all the data elements of the original, the data can be aggregated and dissembled in a variety of configurations according to the ultimate information requirements

3. Much of the data are standardized within and among institutions.

Disadvantages of using existing data sources are:

1. The data may have to represent something they were not intended to represent, e.g. cataloging data may be used to represent materials acquired. Some acquisitions are never cataloged or are delayed in cataloging.

2. The data may not be complete enough to serve its secondary purpose. e.g. many libraries using automated cataloging systems to generate card catalogs do not update holdings statements. Using incomplete data may misrepresent the acquisitions situation.

3. Libraries do not uniformly input all cataloging in their automated cataloging systems. Library policies differ as to the cataloging of government documents, technical reports, non-Roman alphabet materials, materials in other than printed format, etc.

4. Data elements such as the Library of Congress card number and ISSN are not uniformly assigned to all materials in the publishers' records or in library bibliographic files. In addition, some data elements are incorrectly recorded preventing the match which should occur when comparing records.

5. The data are limited by the coverage of the data base. For instance, Book Publishing Record includes only books published in the U.S. or having a sole U. S. distributor. It also covers a limited time period. MARC tapes are limited because of selectivity and language. There is no universal bibliographic data base in suitable machine readable format covering all formats, languages, countries, etc.

6. Data generation is susceptible to human error. The classification of library material is a matter of judgment. Similarly, academic courses and programs are assigned a HEGIS category based on the best judgment of an expert. These decisions are not reviewed for overall consistency among the library and academic universe. They do, in fact, represent local and personal interpretations of subject matter.

In spite of the disadvantages of using existing data, the alternative is to set up separate systems. Libraries cannot support data gathering systems to provide collection development information. There is much, however, that can be done at relatively little expense to improve initial input and make the existing records more useful for

collection development. These adjustments include:

1. Make the machine readable record the master record. Input all holdings changes on this record.

2. Input all data which the library determines will be useful for collection development in the appropriate MARC field. Portions of the MARC 008 field concerning government documents, form of publication, imprint date, type of publication, and bibliographic level, might be especially useful.

3. Review the classification number and subject headings to assure that they accurately represent subject matter.

4. Adopt a library policy to input all new acquisitions within a reasonable period after their receipt. This will ensure that the cataloging record reflects current acquisitions.

5. Adopt a policy to include all cataloging in the machine readable data base regardless of cataloging level. This will ensure that all materials in the collection are represented and will enhance the utility of the data base as a union catalog and a shared cataloging system.

Adoption of these adjustments also, of course, represents a considerable improvement in the libraries catalog and are, in fact, necessary in the world of the on-line catalog (which will be a fact of library life within the next few years). In short, the improvements are necessary for their own sake. The management system can accommodate the improved data.

Data Aggregation.

Applying the data collection management decisions requires aggregation. The rationale for this process is that managers need to weigh data from several sources and cannot cope with unaggregated data. Aggregation permits the manager to discern patterns in the data, to compare one set of data to another and to predict future outcomes.

In collection management, the primary aggregations are suggested by previous research and collection development methodology.

Aggregation by Subject.

To determine how to aggregate the data elements listed in Table I, academic libraries can look to university management information systems. Federal, higher education statistical reporting utilizes subject content as its major categorization. In order to determine the numbers of people trained in various subjects and career categories, and to measure the focus of American higher education generally, the U.S. Office of Education requires the reporting of enrollments according to a standard taxonomy known as the Higher Education General Information Survey (HEGIS) Taxonomy. (Huff, 1970) Some state governments have extended HEGIS reporting to meet their own management requirements, and some universities, including SUNY, use HEGIS as the basis for their management information systems. A standard taxonomy of disciplines is necessary to discern trends in manpower training and make projections regarding educational trends.

Library materials, too, are classified by subject. Classification numbers and subject headings are assigned to each book during the cataloging process. Collection development is based, to some extent, on the subject content of the materials. Libraries define collecting profiles based on broad subject categories or on subject heading groups or segments of the library classification scheme.

Additionally, publishing output as reported in sources such as Book Publishing Record (BPR) carries LC class numbers and subject headings and has been subject categorized using a set of broad subject categories related to the Dewey Decimal Classification.

A further argument in favor of the use of subject aggregations is that previous attempts to model aspects of the collection development/collection use process have concluded that subject content is a useful mode of analysis. For instance, Kent, et.al. (1977) and Chen (1974) found that the subject distribution of circulation remains relatively stable over time. McGrath (1975) and Pierce (1976) found that variations in circulation by subject can be predicted, to some degree, by, variations in student enrollment, subject characteristics, and level of student.

Most academic libraries take the view that their collections should reflect the teaching and research programs of the institution. This implies a relationship between the subject matter of the collection and the subject matter of the curriculum and research programs. Collection development policies are typically organized according to subjects and delineate the extent to which each subject shall be collected. Collection development budgets are frequently allocated by subject and represent an attempt to plan the subject distribution of current acquisitions.

Much of the recent research in collection development has analyzed subjects within a collection. In addition to the Kent, Chen, McGrath and Pierce studies cited earlier, a number of studies have used subject as the unit of analysis. Knightly, (1975) and Raper, (1969) both attempted to compare current acquisitions with institutional curriculum. One is concerned that libraries control for their current acquisitions by subject, lest inflationary cost increases in some subjects unduly diminish collections in other subjects.

The lack of subject analysis in other studies is also instructive. Some studies of user satisfaction and book availability, (Buckland, 1975; Saracevic, Kantor, Shaw, 1977) fail to provide subject breakdown of satisfaction rate. Subject analysis would guide selectors in determining which subjects are providing satisfactory service and which require improvement. A recent study of document delivery capability conducted in Australia provides an analysis of six rather broad subject areas, and demonstrates the diagnostic advantage subject analysis gives in detecting collection strengths and weaknesses. (Wainwright and Dean, 1977)

A further indication that analysis by subject is an accepted practice in collection development is the recommendation of the American Library Association, Collection Development Committee that libraries adopt a cut of the Library of Congress Classification to define their collections and report collection statistics thus permitting inter library comparison. (Guidelines, 1977)

Finally, academic library collection standards and formulae are often based on breakdowns with regard to the level at which subjects are offered. The Clapp-Jordan and Washington State formulas (1965, 1970) both suggest a certain number of volumes per baccalaureate major. Increments for advanced degrees in a subject or student

enrollments exceeding a stated figure are specified. Voigt (1976) also offers differing levels of acquisition for a variety of Ph.D. subjects. Even though the empirical justification for these and similar formulae are in question the concept of planning and buying library collections to support academic and research programs of the parent institution is deeply ingrained in collection development theory and practice. After all, that is the primary purpose of an academic library.

Subjective collection analysis methods such as the comparison of library collections with accepted, selective or comprehensive bibliographies are also subject oriented (Lopez, 1972). Edelman, (1977) suggests that disciplines vary greatly in the proportion of new versus retrospective materials, primary versus secondary material, periodicals versus books, reference and bibliographic sources, and written texts versus other recorded formats. This suggests that further aggregations within subjects may also be useful.

Selection of Subject Groupings.

The problem is to assay the needs of an academic library based upon knowledge of the academic program it supports. Since the link between academic programs and libraries is discipline or subject matter, it appears that this is the appropriate unit by which to measure and predict library needs. Since no existing taxonomy describing either academic programs or library materials is acceptable to all potential users, and no taxonomy is available which unites library materials and academic programs, the problem is to locate the best available system for each purpose and provide a translation between the two.

With the exception of the ALA Collection Development Committee's Guidelines, defining a segmentation of the Library of Congress Classification (1976) and the Berkeley study on which this segmentation is based, there has been little effort to define subjects for library collection analysis. Libraries typically define their collection profiles in terms of academic departments or in terms of a set of locally defined subjects. The computer programs used in this project are capable of handling either CONTENT or COMPONENT subject analysis. Both are described in the final report of a previous project by this agency (Evans, et al 1977.)

If managers require comparisons between library collections, publishing output, acquisitions and academic and research programs, then both bibliographic records and academic records should be classified according to selected subject groupings. The disadvantage of establishing a set of subjects unique to a given campus is primarily that comparisons cannot be made among similar libraries, or among libraries organized for resource sharing such as members of a consortium or individual units of a multi-campus university.

The prospect of identifying a set of subjects to accommodate all the diverse academic and research activities of a broad spectrum of institutions as well as library collections at these institutions posed a challenge. The first step was to identify the various classifications in use with academic programs and library collections, select the most appropriate one for each and design a mode of relating one to the other.

Classification of Academic Programs.

Similarly named departments and programs have different emphases at various campuses; thus it is desirable to select a disciplinary taxonomy specific enough to identify these differences yet broad enough to identify programs sharing common elements. The taxonomy must be understood and accepted by a sizeable group of higher

education institutions to provide a sufficient data base and to assure the MIS will be useful to a range of institutions. It should be an established taxonomy so that longitudinal comparisons can be made.

The most widely accepted taxonomy for higher education programs is the Higher Education General Information Survey (HEGIS) (Huff, 1970) which is used for Federal reporting of programs and for campus administrative data. Some states also use HEGIS for academic program registration. Although this system was developed for reporting enrollees and graduates of various programs, it has been adapted in some institutions to assess academic management data, e.g. cost per student credit hour, or faculty/student ratio, by the subject content of courses. Many institutions including SUNY report subject content using the HEGIS Taxonomy. Considerable investment has been made in the HEGIS system. The National Center for Educational Statistics (NCES), which controls the format of HEGIS, was preparing to contract its revision at the start of this project. An interview with the Director of Surveys at NCES indicated that the revision would resemble the present HEGIS in form and structure but would incorporate academic programs instituted since the original HEGIS was published in 1970.

The SUNY Course and Section Analysis (CASA) system classifies all SUNY courses according to subject content using a modified version of HEGIS (See Appendix 1).

Subject Content Versus Program Classification.

The stated purpose of the HEGIS Taxonomy is to describe academic programs and quantify the numbers of persons receiving training in these programs. A secondary purpose which is evident in the SUNY Course and Section Analysis files is that of determining the academic activities of a university by subject. All SUNY courses are assigned a HEGIS number based on their subject content rather than on the department or program of which they are a part. For example, courses offered by a speech department may be assigned to HEGIS 0812, Education of the Deaf; HEGIS 0896, Speech Education; HEGIS 1220, Speech Pathology and Audiology, and HEGIS 1506, Speech, Debate, Forensic. This assures that courses with similar subject content are assigned similar HEGIS numbers.

The work on the LC/HEGIS Translation was undertaken based on the SUNY use of HEGIS, i.e. the categories represent subject content rather than programs. This interpretation is easily maintained in sections of HEGIS: law, engineering, letters, and social sciences can be interpreted either as programs or subject matter. Health professions, (HEGIS 12), present a problem because many of these categories are clearly not subjects, e.g. Dentistry. If one were to interpret these as programs, one would need to include much of HEGIS 04, Biological Sciences, which are basic to the health professions. Also, the basic medical literature would have to be repeated in each of the medical professions. This is one of the reasons for omitting HEGIS 12 from the translation as will be detailed in a later section.

The 5000 and later sections of HEGIS are the sections dealing with pre-baccalaureate technology subjects. These, too, are designed purely with an emphasis on describing programs rather than subject content. The approach on these categories has been to establish the content of each program and then to assign the LC ranges and numbers which apply to the subject content of the programs, e.g. HEGIS 5202 Dental Assistant Technologies would include general biology, anatomy, physiology, radiology, business mathematics, office management, and all of the other subject content included in a program in dental assistant technologies as well as the specialized work in anatomy of the mouth and dental technology. This procedure necessitates much overlap with

other sections of the Translation. Biology will be repeated in almost every category of 5200 Health Services and Paramedical Technologies.

Comparison of HEGIS with SUNY Modified HEGIS.

The two taxonomies are essentially the same except that the SUNY version has been expanded to accommodate SUNY programs. These additions are found in the 90's in each discipline and are seen as an expansion of the "xx99 others (specify)". The one major category which has been omitted from the SUNY version is Theology 23. Forestry 24 has been initiated in the SUNY version, and Religious studies, 1510 includes theology whereas in the original HEGIS, Theology is excluded from 1510. The SUNY version of HEGIS is the one used in this project since it is the classification used for all SUNY academic data. It can easily be adapted for more general use by aggregating all of the specific xx9x categories into "other" and grouping forestry subjects into 0114 through 0117. Theology would require expansion of the present 1510 Religious Studies.

Educational Management Service's Educational Subject Matter Classification (Chismore, 1978) has also been examined since it is a system oriented to subject matter as opposed to programs. The subject approach appears to be more compatible with library materials, than the program approach. The ESMC was considered because it is more detailed than HEGIS and incorporates scope notes which aid the classifier in assigning courses to the subject matter.

Study showed that ESMC does not accommodate all academic subjects or all significant library holdings. The greatest disadvantage to using ESMC is that neither academic program information, library collection nor publishing data are available in ESMC and a three-way translation involving HEGIS, ESMC and the library classification would be required.

An international classification of academic programs was recently developed which may have important implications for the future, but it has not been adopted in this country. This is the International Standard Classification of Education (1976). The major problem in using this classification is that no data are available in this system, and a three-way translation is required. Three-way translation is not only time-consuming and costly but is more susceptible to errors of interpretation. The definitions and scope notes in both ESMC and the International Standard Classification of Education are extremely useful in the present project even though the SUNY modification of HEGIS is the classification scheme selected for conversion in this project.

Classification of Library Materials.

There are three common systems which are used to describe subject content of bibliographic records. These are: Dewey Decimal Classification, Library of Congress Classification (LC) and Library of Congress Subject Headings. National Library of Medicine Classification (NLM) and Medical Subject Headings (MESH) are used for medical collections. Most academic libraries describe the subject content of their collections in one or more of the above taxonomies, though Government documents, technical reports, non-print media, and periodicals are often not classified.

Machine readable commercial data bases, such as R. R. Bowker's Books in Print and Book Publishing Record and F. W. Faxon's data base of current periodical titles, also contain subject information, notably LC and Dewey Class numbers.

A fourth potentially useful subject classification system has recently been developed. The Broad Subject Ordering System (BSO, 1977) was developed as a

translation between various library and bibliographic classifications for use in bibliographies and other publications for which broad subject classification is required. Unlike the systems mentioned above, BSO has the capability of faceting, i.e. expressing the relationship of two subjects in a single notation. An alphabetical index is provided, and UNESCO is committed to updating and revising the system as required. This system offers considerable promise in the future especially if the American library profession encourages its inclusion in new MARC records or provides a translation between present library classification systems and BSO. The level of detail and the terminology employed in BSO was useful in the current project though BSO was not the primary classification used.

PRECIS (1969), a faceted system of subject headings incorporated in the British MARC and British National Bibliography records, is gaining interest in this country. PRECIS offers potential utility in classifying bibliographic records,but there are not enough PRECIS records available at present to warrant its use in this project.

The LC classification number is the most prevalent single expression of the subject content of academic library materials. The LC number has a limited length, uniform format and is easily manipulated by the computer. LC subject headings, on the other hand, are long, contain punctuation and are assigned less rigorously by libraries. Each bibliographic record may have several subject headings but only one classification number. Dewey is less prevalent in academic libraries and in MARC records, though it might be preferred for a public library MIS.

The selection of a subject classification is influenced by the sheer weight of investment in LC and HEGIS which places alternative taxonomies at a grave disadvantage even though they may display considerable theoretical and practical advantages. It was decided, therefore, to use HEGIS and LC Classification as the two taxonomic systems to relate academic programs to library materials.

This project is not the only one to recognize the value of a cross-over table between HEGIS and LC. The University of Colorado adopted this approach in planning an internal management information system in 1977. The result of this work was a much less detailed translation. (Colorado, 1977)

This section has described the need to aggregate data by subject in order to make comparisons between academic programs and library materials. Subject aggregations are indicated by collection development theory and practice. Possible methods of aggregating bibliographic and academic data by subject were discussed, and rationale for designing a translation between LC and HEGIS presented. The following section presents the methodology used in the construction of the LC/HEGIS Translation, progress in its completion and factors critical to interpreting the reports generated using the LC/HEGIS Translation.

4. LC/HEGIS TRANSLATION

Methodology of Developing a Translation between HEGIS and LC.

Once HEGIS and LC were selected as the appropriate classifications, a translation was needed to relate the two so that bibliographic data and academic program data could be correlated. Several methods were explored to construct the translation hereafter referred to as the LC/HEGIS Translation:

1. Using subject specialists to describe the HEGIS categories (instructional subjects) via LC numbers. (subjective)

2. Classifying library materials by the academic program affiliation of individuals who use them. (empirical)

3. Surveying library users as to their perceived library requirements. (subjective and empirical)

4. Clustering LC classifications of documents which have co-occurring subject descriptors. (empirical)

Subject specialists were employed in the translation construction due, in part, to the absence of available data for options two and three and lack of conclusive research of the techniques and validity of automatic classification (option four) in a multidisciplinary data base.

The subject specialists were selected primarily for their expertise in cataloging or in providing information services in defined subject areas. Many of the specialists work in special libraries and are aware of the information needs of persons working in the subject area. A list of the subject specialists and the areas they contributed is reported in Appendix 2. The general methology and workflow of the LC/HEGIS Translation are presented in Figure 4-1

Several practical and theoretical problems emerged in the attempt to relate the two classifications:

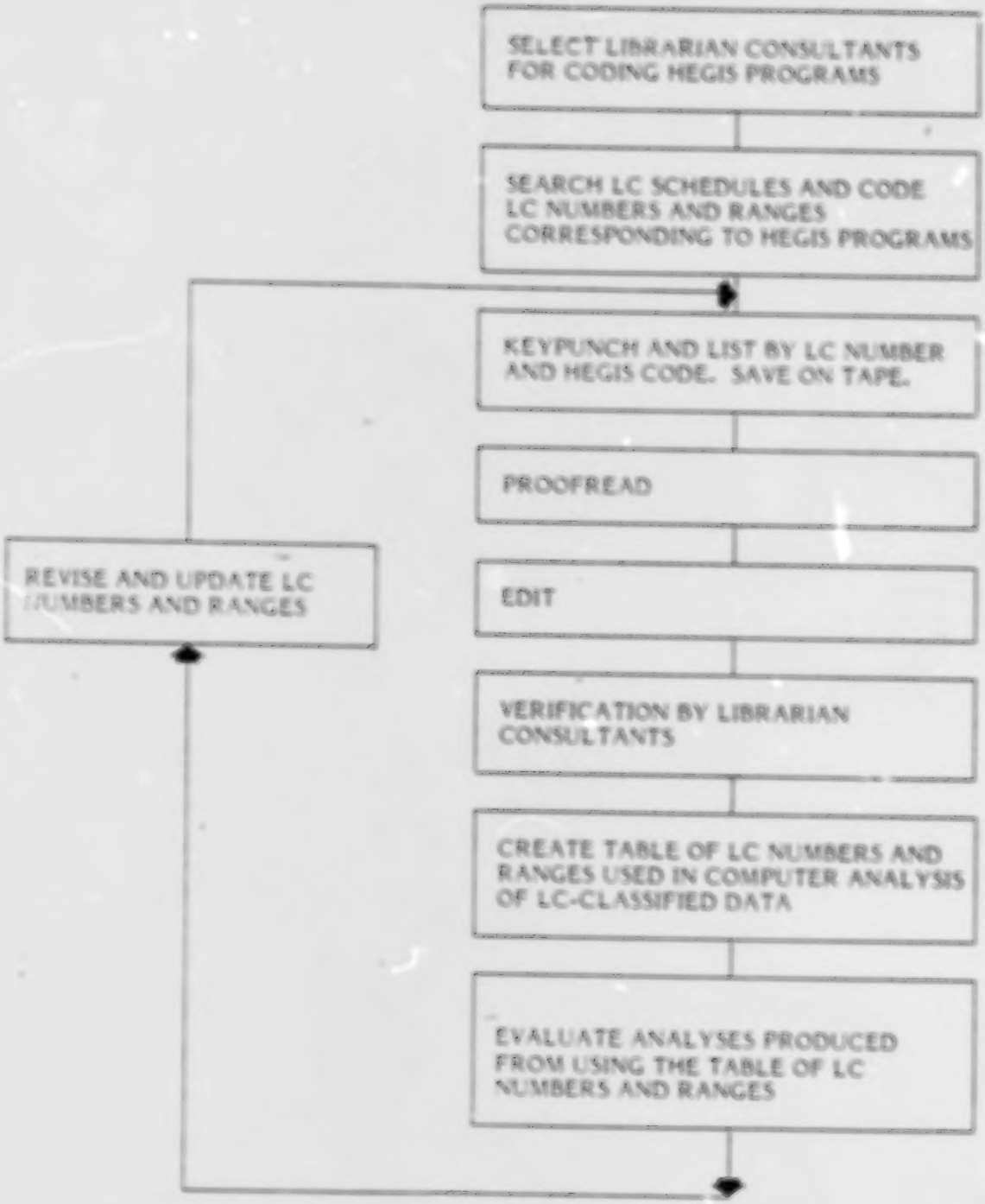
1. The two classifications have incongruent structures.

2. Subject matter is not discrete.

3. Knowledge and knowledge classification schemes are volatile.

Figure 4-1

HEGIS/LC TRANSLATION WORK FLOW


SELECT LIBRARIAN CONSULTANTS FOR CODING HEGIS PROGRAMS
SEARCH LC SCHEDULES AND CODE LC NUMBERS AND RANGES CORRESPONDING TO HEGIS PROGRAMS
KEYPUNCH AND LIST BY LC NUMBER AND HEGIS CODE. SAVE ON TAPE.
PROOFREAD
EDIT
VERIFICATION BY LIBRARIAN CONSULTANTS
CREATE TABLE OF LC NUMBERS AND RANGES USED IN COMPUTER ANALYSIS OF LC-CLASSIFIED DATA
EVALUATE ANALYSES PRODUCED FROM USING THE TABLE OF LC NUMBERS AND RANGES
REVISE AND UPDATE LC NUMBERS AND RANGES

Incongruent structures.

HEGIS consists of a two level hierarchy of subjects. The superior level is defined by a set of two digit categories. The subordinate level is defined by a set of four digit categories the first two of which are the two digits of the superior group. (See HEGIS taxonomy Appendix 1) Botany, Plant Pathology and Zoology are represented as equivalent subordinate categories under Biological Sciences; whereas traditional classification practice would place Plant Pathology in a tertiary level under Botany.

The Library of Congress Classification consists of an unspecified number of levels. Each subject is subdivided as many times as required in the classification of library materials. Divisions are largely based on the actual books classified at the Library of Congress. Whenever it appears that a new breakdown is required and the notation permits, the classification is revised.

Idiosyncrasies and Complexities of the Classifications.

Both LC and HEGIS are uneven in level of detail. For example, HEGIS divides psychology into more than 20 categories; whereas sociology, political science and history are allocated one category each. In HEGIS, it is impossible to discriminate between European and American history. The detail of the translation is such that additional subject breakdowns within each HEGIS category can be provided. (See Appendix 1) These subdivisions will be used for collection analysis, but no attempt will be made to subdivide academic programs since this would require reclassification of courses and an expansion of HEGIS.

Pseudo HEGIS categories.

The HEGIS Taxonomy provides for expansion in the final portion of each two digit category. Several expansions have been made in the SUNY-modified HEGIS. In the course of creating the Translation, a few subjects were identified which required separate categories which did not coincide with the established HEGIS categories. Assigning regular HEGIS numbers was not possible, due to possible conflict with future HEGIS assignments. Therefore, pseudo HEGIS numbers were generated to retain the identity of these subjects and the ranges associated with them:

0599P	Production
0899R	Recreation Education
1599F	Fiction (Contains fiction classed as PZ)
1599x	Other Literature

Additional pseudo HEGIS numbers will appear in the technology subjects.

Although there are obviously no enrollment data with which to match these pseudo-categories, the new categories are valuable to both library and academic administrator as expressions of library and publishing data.

Uneven growth of the LC Classification dictates the need to specify up to two decimal points and twenty-five total positions in the classification number in order to identify some of the ranges and numbers relevant to instructional subjects. For example, Nigeria has one whole number and several decimal positions to describe all aspects of its history and culture, whereas Poland has several whole numbers and decimals.

Due to discrepancies in the classification schemes, one classification has to be fitted to the other. Since HEGIS has the simpler structure LC must be fitted to HEGIS. The basic unit of the translation is the four digit HEGIS number. The subject specialists were assigned one or more four-digit HEGIS categories and asked to locate and list all LC numbers and ranges relevant to each four digit HEGIS category. Specific instructions regarding the syntax and coding requirements were developed and issued. (See Appendix 3)

Operational Matters.

Progress on the Translation. Assigning LC numbers and ranges to HEGIS four-digit categories results in a list of some 175 four digit HEGIS categories mapped against approximately 12,000 LC ranges and numbers. The translation work focused primarily on baccalaureate subjects taught at the 22 SUNY state operated campuses. Health professions (HEGIS 12) was partially omitted for several reasons:

1. The HEGIS categories describe professional specialities rather than subject matter.

2. Our medical library consultant advises that the specialties offered have little influence on collection patterns of medical libraries.

3. Medical libraries typically use the National Library of Medicine Classification rather than LC.

4. Virginia Bowden at the University of Texas Health Sciences Center Library in San Antonio is working on a CLR-funded project to develop a method of analyzing Medical library collections using NLM Classification. The computing procedures established by the SUNY project can accommodate this classification through the definition of new tables and parsing algorithms.

For these reasons, it was decided to defer the majority of health profession HEGIS categories and concentrate on other baccalaureate categories. A few health categories such as speech pathology, physical therapy, first aid, and public health are offered at the baccalaureate level in a number of SUNY units and are thus included.

Technology and pre-baccalaureate subjects were deferred until the completion of the baccalaureate subjects because it was felt that the former would be heavily dependent on the latter. Technical subject specialists based their work on the ranges established by subject specialists working on the baccalaureate subjects. The work on the technology subjects was completed by the end of the project, but because of the lead time required to do the analyses, these were not included in the version of the Translation used for the analyses.

Area studies were contracted near the end of the project due to the perceived difficulty of defining area studies. It was thought that experience gained in translating the basic baccalaureate subjects would be useful in doing the interdisciplinary area studies. Again, due to the early cut-off date required to run the analyses, some area studies are not included. Area studies considered too immense to tackle within the operational confines of this project (e.g. American Studies) have been delayed for later consideration.

Proof-reading. Following coding by the subject specialists, the ranges were keypunched and listed by instructional subject and by LC number to facilitate checking

and editing. Instructions for proof-reading are given in Appendix 6.

At least two concerns emerged in the translation: operational and interpretive. The operational matters relate to the construction, maintenance and use of the translation. Interpretive problems relate to the definition of an instructional subject, mapping the relevant LC ranges, and accommodating overlap among subjects.

Table Construction and Modification. The ranges associated with a four digit HEGIS category are incorporated into totally flexible machine readable tables. Tables are constructed from the individual ranges to represent any individual subject or group of subjects required for a particular analysis. A library may sort its records by individual LC range, by groups of ranges, by four-digit HEGIS grouping, or any combination of the above. Any range or combination of ranges within a four-digit HEGIS category can be modified by simply coding the change and submitting a routine order through the computer center to alter the tables. No alternations need be made in the basic programs. Appendix 5 records an example of a translation table.

Use of the Translation. It is anticipated that libraries wishing to use the translation will define the purpose of the analyses to be done and work with a person familiar with the translation to accomplish their goal. If a comparison is to be made between academic programs and library holdings, then a set of four digit HEGIS categories will form the basis of the analysis. For a SUNY campus, the selection of four digit HEGIS categories may be derived directly from the reports which the campus files with the Office of Policy Analysis or with the U. S. Center for Educational Statistics. In fact, the capability exists to select only the LC call numbers and ranges associated with the HEGIS categories reported by the campus. In using this, however, one must study both the local assignment of courses to HEGIS categories and the construction of the LC/HEGIS translation to be sure the two correspond. There is a tendency for some campuses to report the majority of courses in the general categories of HEGIS rather than in the specific four digit HEGIS categories. For instance, a campus might report 5000 students enrolled in General Psychology, 1000 in Educational Psychology, and none in Personality, Developmental Psychology, or Tests and Measurement. The fact that General Psychology in the LC/HEGIS translation includes only general periodicals and history and theory of psychology (the initial numbers in the BF schedule) should prompt a library to compare its holdings with not only the actual categories in the campus CASA report but also any other specific subjects which are included in the campus course offerings.

Use of the Translation outside SUNY. The tables will provide valuable analyses of library collections regardless of the type of library and the state of local academic management data. Some difficulties may be encountered when attempting to match library data with academic data in some institutions. SUNY appears to be one of the few institutions generating detailed academic information aggregated by subject content. All SUNY courses are assigned a HEGIS category based on their subject content rather than on the department in which they are taught. Assignment at the course level allows faculty assignments, course enrollment, space allocation, etc. to be measured according to subject content.

Institutions reporting majors by academic program lack information on subject content which is desired for library planning. These institutions can take advantage of a secondary analysis tool, the Induced Course Load Matrix (ICLM) (Hussain, K. M., 1971). The ICLM permits a breakdown of the subjects taken by majors in a given program through an analysis of their transcripts.

The ICLM matrix for majors in five subjects might resemble that presented in the following example.

SAMPLE INDUCED COURSE LOAD MATRIX

STUDENT MAJORS	SUBJECTS TAKEN									
	Psychology		Sociology		Biology		Literature		Education	
	SCH*		SCH		SCH		SCH		SCH	
Psychology	60	50%	24	20%	18	15%	12	10%	6	05%
Sociology	36	30%	60	50%	12	10%	6	05%	6	05%
Biology	12	10%	12	10%	60	50%	18	15%	18	15%
Literature	6	05%	6	05%	12	10%	84	70%	12	10%
Education	30	25%	30	25%	12	10%	12	10%	36	30%
	144		132		114		132		78	

*SCH = Student Credit Hours

Knowing the number of majors in each program allows an estimate of the enrollment in associated-subjects. The ICLM has been found to be relatively stable over time and provides a relatively reliable and inexpensive method of estimating subject enrollment without classifying each course and tracking all the associated data. It permits institutions having only program data to convert it to subject content data.

Use by Non-SUNY Libraries. Non-SUNY libraries wishing to use the MIS to relate institutional variables with library variables will provide their own institutional data. If the data is not classified according to the HEGIS taxonomy, either a correspondence table will have to be provided so that the LC/HEGIS translation can be related to the institutional data or a new translation must be undertaken. Institutions which report data only by departmental major may consider developing ICLM matrices to describe the overlap between departments and the impact of non-majors on the demand for a given subject.

Institutions which map individual courses to HEGIS according to subject content may use the LC/HEGIS translation directly.

Analyses of Bibliographic Data only. It is possible to analyze library collections without any reference to the institutional variables. In these cases the library or group of libraries may construct a different set of instructional subjects for analysis of their bibliographic data. These analyses need not be tied to the LC/HEGIS translation but there is considerable value in the development of comparable data, and the use of the existing tables (the building and maintenance of which are not trivial tasks) would avoid additional expense. In the case of a public library, one would expect that analyses would be performed on general classes, with probably special concern for local interests such as commerce and industry, local and regional history, as well as on general vocational and recreational requirements.

Updating and Revising SUNY LC/HEGIS Translation

Since knowledge classification systems grow and change in accordance with the growth of knowledge, the translation must accommodate growth as well. The translation will be revised by periodically reviewing additions and changes to the classifications.

A fundamental premise of the system is that each class number represents a physical item stored in a library. Feedback from book shelf to data analysis through a review of titles included in the analysis and identification of known, relevant items excluded from analysis because of omissions in the translation will be sought. Both cases suggest revision of the tables to include or exclude specific ranges. Specific suggestions regarding revision of the Translation are found in the final section of this report.

Interpretive matters

Overlap. Overlap among instructional subjects presents an interpretive problem. The decision was made to confront the issue of overlap and allow LC numbers to be assigned to more than one instructional subject. The rationale is that library materials may support work in more than one subject.

Two instructional subjects are said to "overlap" when the same class number or range of class numbers appears in each disciplinary table, and hence are relevant to both of them. In a practical sense as a noun, the word overlap refers to all titles found in the analyses that apply in common to more than one instructional subject. In a more general sense, the collections of two libraries may be said to overlap when they contain the same titles; hence the term overlap is not limited to classification by subject.

Previous attempts to analyze library collections by subject have ignored overlap by defining subjects discretely. However, overlap is a central issue to library collection management since, for example, if a course or department is dissolved, only these parts of the library collection relevant to that department and no other are undisputable candidates for reduction. This is because much of a library collection supports more than one instructional subject; that is, overlap is the rule rather than the exception.

Overlap may present serious problems of interpretation and administration. If the same title applies to more than one subject, then in a budgeting sense, who should pay for it? There are several possibilities.

1. Allocate one copy to each subject which it supports.

2. Allocate the title equally among the subjects.

3. Divide the allocation for title among the subjects according to some criteria, such as projected use in each subject.

No decision as to the best alternative has been made in this project, nor should it be. Such a decision should be left to the policy makers of the institution or the library. From the standpoint of value, however, we suggest that overlapped titles be computed as generating multiple values rather than one value to be divided among departments.

The solution may vary according to the type of data and the purpose for the analysis. If a library wishes to identify all of the books in its collection which are relevant to a projected, new graduate program it would be important to count all of the

relevant books regardless of their overlap with other instructional subjects. When estimating money needed to buy relevant books from the current publishing output, one would not fund the same book in several instructional subjects unless it were demonstrated that multiple copies were needed.

The programs developed thus far permit only the first alternative: that of attributing titles to each instructional subject they support. This solution was judged to be the most generally applicable to collection development, but some problems in applying this solution to costs will be brought out in a later section. When components are aggregated, the aggregate counts each title once no matter how many subcomponents contain it. Thus, the sum of titles in subcomponents will exceed the number of titles in the aggregate if overlap exists between two or more subcomponents.

EXAMPLE OF AGGREGATE COUNTING

OF OVERLAPPED TITLES

SUNY GROUPING OF HEGIS PROGRAMS	4696
EDUCATION	455
MATH AND COMPUTER SCIENCE	369
PHYSICAL SCIENCE AND ENGINEERING	811
PSYCHOLOGY	1646
SOCIAL SCIENCES	3066
NUMBER OF TITLES ANALYZED	4696
NUMBER OF ATTRIBUTIONS MADE TO HEGIS CLASSES (455 + 369 + 811 + 1646 + 3066)	6347
ADDITIONAL ATTRIBUTIONS	1651

In order to project purchases by instructional subject, considering that material bought for one instructional subject may very well support other subjects, one must aggregate the instructional subjects totally to squeeze out all duplication of individual titles. This is not a difficult process using the computer programs described earlier in this section. Some would argue that items serving more than one subject should be purchased in multiple copies. This opinion may be verified empirically by analyzing use data to determine whether material which supports several instructional subjects accounts for larger proportions of circulation than material which supports few subjects. A model could be constructed to define the amount of duplication necessary in each instructional subject or among instructional subjects to meet the needs of users within a given setting.

Terminology. At this point it is necessary to present terminology used in the reports. Since a single item may be counted or listed within several subject categories it is necessary to define what that item's presence means. The term "support" is used to indicate that the classification number of an item falls within an LC range mapped to a four-digit HEGIS category or a combination of four-digit HEGIS categories. It is assumed that the subject specialists believed that any item classified in the ranges selected to define a four digit HEGIS category in some way "supported" work in that subject.

The term "instructional subject" refers to a four digit HEGIS category or any aggregation of four digit categories along with the LC ranges mapped to these HEGIS categories. The instructional subject serves as a unit of analysis.

Serials. The question of whether the translation tables could be used to sort serial records was discussed. The major available serial data base having LC class numbers is the Faxon data base. In that data base, the LC numbers consist of only the first two letters of the class number. The translation is too detailed to handle this data base. However, it can handle adequately a set of serial records containing the complete LC class number. This was demonstrated in running an OCLC tape containing 1200 CONSER serial records against the translation tables and printing out the serial titles by four-digit HEGIS category. The titles displayed, in almost every case, corresponded to the categories. The few exceptions indicate areas of the translation requiring modification. We therefore conclude that the translation is valuable for the analysis of serials collections.

Summary

This section presented the methodology involved in constructing the LC/HEGIS translation and the use of the translation in aggregating data by subject. The tables representing the translation are the primary basis of subject Content Analysis in the management information system. The translation provides a link between academic programs and library collections as well as a means of subject analyzing a single collection or group of collections. The LC/HEGIS translation is only one set of tables which could be constructed for use in the MIS. Other tables which are potentially useful are:

1. Translation between LC and Dewey Classifications.

2. Translation between LC Classification and LC Subject Headings.

3. Definition of additional levels in the current LC/HEGIS Translation—aggregation of individual ranges into several levels.

4. Translation between Universal Decimal Classification and HEGIS.

5. THE PROJECT FROM A DATA PROCESSING PERSPECTIVE

This appendix is addressed to the data processing staff member who is charged with carrying out similar analyses for libraries. The term "user" is used to refer to the librarian, or administrator for whom the analyses are carried out.

Design Objectives

The primary program design objective from the start of the first project in 1976 was flexibility. This was because the programs were to support research in collection development analysis and, therefore, be quickly adapted to new directions. Secondly, flexible programs can be made more efficient through experience either by eliminating some unused capabilities or by reprogramming the time consuming sections.

Flexibility has manifested itself in two major ways in the design of the programs. First, the programs are table-driven, and second, each is largely independent of input and output.

Tables

There are several types of tables serving different purposes. Field definition tables define the position, length, format and function of data elements that are both input and output. Parse tables define how a data element (e.g. LC call number) is to be transformed into a standard format. Hierarchy tables describe the components of the hierarchical system being analyzed (this might be the hierarchical breakdown of a library collection by LC class, or the hierarchical structure of campuses in a university). The hierarchy concept is an important one and its application in this project will be explained more fully later in this section.

Although each type of table has a different format, all tables are input on card-image records so that the same editing facilities that a programmer uses to create and update programs may be used to create and update tables.

Except for hierarchy tables, all tables are input free-format using the data-directed input technique of the PL/I programming language:

```
START = 1,  LENGTH = 4,  FORMAT = 'CHAR',  FUNCTION = 'COUNT';
```

This example might be the description of an input or output field. Each field parameter is coded as a key word, followed by an equal sign, followed by the value of the parameter. Parameters describing one table entry (in this case, a field) are coded consecutively, separated by commas, and ended with a semi-colon.

Hierarchy tables are not input this way because they are so large. For example, the hierarchy tables defining the HEGIS/LC translation contains over 12,000 card images.

I/O Independence

To say that the programs are largely independent of input and output means that the input and output fields are described by parameters in a field definitions table. An LC call number, for example, could be taken from anyplace in the input record, parsed, and then placed anywhere in the output record. This enables the user to utilize existing data without writing intermediate programs to re-format the data.

While I/O independence is an established and popular programming practice, it has its drawbacks. First, programs do not operate as efficiently because the position and length of data fields are variable. A second drawback is that it takes more time to set a job up for execution since the record layout must be redefined each time.

In spite of the difficulties, I/O independence has proven to be very useful. It enables the same programs to process OCLC distribution tapes; R. R. Bowker, BPR data; and SUNY CASA data simply by defining different tables.

The Hierarchy Concept

A second major design decision was that the capability to analyze the library collection should be entirely general as to the unit of analysis (LC class, HEGIS category, faculty research area, special collection) and the levels of aggregation. This generality is more precisely described in terms of the following principles:

1. It should be possible to analyze the library collection using any data element in the OCLC/MARC format or in other formats in which the position and length of data elements are fixed.

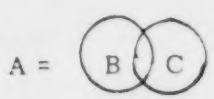
2. It should be possible to translate the data element as expressed in the input file to a data element that is more useful to the user. In the current project, it was necessary to translate LC to HEGIS; that is, to express each four digit HEGIS code as one or more LC call numbers or ranges. It was also necessary to translate the holding institution codes used by OCLC to the campus codes used by SUNY.

3. It should be possible to aggregate the units of analysis (e.g. four digit HEGIS categories, individual SUNY campuses) into larger units of analysis (e.g. two digit HEGIS categories, four year colleges as a group) in any number of levels (individual SUNY campuses, four year SUNY college, SUNY, public universities in the U.S., all universities in the U.S.)

4. It should be possible for units of analysis to overlap (e.g. all universities in New York State, public universities in the U.S.). Overlap might occur even in the translation where the same LC class number, for example, might be assigned to more than one four digit HEGIS code due to its general applicability.

In order to satisfy these principles, a new data structure was defined called a hierarchy. A hierarchy is composed of classes and an explicit recording of the relationships among classes. The contents of each class is made of the contents of its subclasses whose contents are similarly defined by their subclasses.

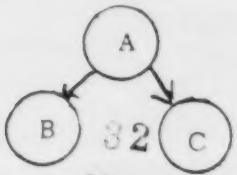
VENN DIAGRAM MODEL



BOOLEAN LOGIC MODEL

A = B OR C

HIERARCHY MODEL



A, B, and C are classes in the hierarchy which might correspond to (Biology), (Botany), and (Zoology) respectively.

In an intuitive sense, what this says is that the entity (Biology) has attributed to it whatever (Botany) and (Zoology) have. If Botany has a collection of 15,000 titles and Zoology has 5,000 titles, then (Biology) has at most 20,000 titles attributed to it.

Note that there is no statement as to the overlap that might exist between B and C.

Very likely some titles might be applicable to both Botany and Zoology. Suppose 2,000 titles are held in common. Then the number of unique titles contributed to Biology is 19,000, not 20,000. Actually, both ninteen and twenty thousand are useful numbers with respect to Biology but must be carefully interpreted. The capability to count the number of unduplicated items in a class is something which these programs have but which most software packages do not have.

Hierarchy Representation. The hierarchy data structure is represented by four hierarchy tables.

1. class descriptors
2. contains relations
3. subclass relations
4. ranges

The class descriptor is the text description of a class, and may be up to 60 characters in length.

Each class is uniquely identified by a class identifier (hereafter called a class id) which is six characters. Since each of the above tables describes something different about each class, the information is linked together by the class id.

The subclass relations table contains, for each class id, the ids of all its subclasses:

Class id	Subclass id
A	B, C

The contains relations table is (almost) the inverse of the subclass relations. It contains the ids of all classes that contain a given class, for each class as follows:

Class id	Containing class id
B	A
C	A

The contains relation table is more than the inverse of the subclass relations table because it contains all containing classes while the subclass relations table contains only direct subclasses. The difference is not apparent until there are at least three levels in the hierarchy as in the following example:

Subclass relations table

Class id	Subclass id
A	B, C
B	D, E
C	E, F
D	-
E	-
F	-

Contains relations table

Class id	Containing class ids
A	-
B	A
C	A
D	A, B
E	A, B, C
F	A, C

Note that A is a containing class for D, but that D is not a subclass of A

The contains relation table is the mechanism by which titles are assigned to each class only once no matter how many subclasses contain the title. The contains relations are a list of classes (along with the class itself) that a title should be assigned to. The fact that no title is counted more than once in a class is because no class appears in the same list more than once.

The subclass relations table enables the contents of each class to be broken down by subclass. When a title is assigned to a class, two fields are output - the class id and the subclass id. If a title belongs to more than one subclass, then an output record is generated for each. The subclass relations table is also the mechanism by which the user builds or modifies the structure of the hierarchy. The contains relations table is generated from this table by the CONTAIN program.

The contains and subclass relations are represented in the computer as a binary matrix, which is a square array of zeros and ones. For the previous example, the matrices would be as follows:

Y / X	A	B	C	D	E	F
A	0	0	0	0	0	0
B	1	0	0	0	0	0
C	1	0	0	0	0	0
D	0	1	0	0	0	0
E	0	1	1	0	0	0
F	0	0	1	0	0	0

SUBCLASSES: X is a subclass of Y

X/Y	A	B	C	D	E	F
A	0	1	1	1	1	1
B	0	0	0	1	1	0
C	0	0	0	0	1	1
D	0	0	0	0	0	0
E	0	0	0	0	0	0
F	0	0	0	0	0	0

CONTAINS: X contains Y

A one (1) at the intersection of a row and column means that the class associated with the row is a "subclass of" or "contains" the class associated with the column. A zero (0) means no relation exists between the two classes. The reader may verify that the matrices contain the same information as displayed by the hierarchy diagram previously shown.

Since most of the matrices are zeroes, they are stored in a compacted form. Each row of the matrix is broken up into segments of eight bits (i.e. binary digits) each, and only segments that are not all zeroes are actually stored. Another number, the position of the segment within the compacted row, is also stored to preserve the relationship between column in the matrix and class.

This representation of the inter-relationships among classes in a hierarchy has proven itself to be quite efficient both in computing time and memory space. In uncompacted form, for example, the hierarchy for HEGIS categories, comprising about 600 classes, would require 90,000 bytes (i.e. one byte is the storage required for one character) to store both the contains and subclass relations. In compacted form, however, they take only 15,000 bytes. Computing time requirements for binary matrix processing is more difficult to ascertain but the compacted form enables both quick scanning of a list of classes and boolean operations lists of classes.

Programs

The software used for the project consist of:

1. PL/I programs written for library collection analysis,
2. Statistical Package for the Social Sciences (SPSS),
3. IBM's Sort/Merge Product,
4. SUNY Central Administration Select Utility.

SPSS is a proprietary statistical package leased from SPSS Inc. of Chicago, Illinois which is used by many universities and businesses.

The SELECT Utility is a very flexible program written by a member of the SUNY Central Administration computing staff. It selects portions of input records using boolean operators. Although other software packages were also available on the SUNY Central computer and would have expedited programming (in particular, Informatic's Mark IV report generator), an effort was made to limit the variety of software that would be needed by other institutions who might acquire the programs to perform similar analyses.

Programming for the project enhanced and added to the programs produced for the previous project. A total of twenty-two programs have been implemented. Of these, ten programs are generalized software tools which perform functions required in a library collection analysis system, and the remaining twelve programs are ad hoc solutions to specific problems. Only the general programs are described below.

Extract

This table-driven program extracts fields, subfields, portions of fields and subfields, or combinations of these from OCLC/MARC records. This program has not

been modified from its original form as used for the previous project. Its code was used, however, as the basis for writing the INITIAL EXTRACT program.

Initial Extract

This program was written in response to experience with the EXTRACT program, which indicated that a more efficient extraction program was needed. This program extracts only certain data elements and creates a standard extract record with the following format:

Field		Start	Length	OCLC Originating Field
1.	OCLC record number	1	7	tag 001
2.	holding library code	8	4	tag 049 subfield a
3.	record status	12	1	leader, sixth character
4.	type of record	13	1	leader, seventh character
5.	bibliographic level	14	1	leader
6.	transaction type	15	1	leader
7.	entire 008 field	16	40	field 008
8.	record sequence number	46	8	sequential number of the first block of data of the OCLC record on the distribution tape
9.	LC call number	54	50	first fifty characters of field 090 if present, otherwise first fifty characters of field 050 if present.
10.	Dewey Decimal call no.	94	50	first fifty characters of 092 field if present, otherwise first fifty characters of field 082 if present.

The title field is formed from the 1xx and 2xx MARC fields according to the following rules:

1. If the bibliographic record is a serial and the key title (field 222) is present, use the first 100 characters of the key title.

2. If the bibliographic record is not a serial, or the record is a serial but the key title is not present, use rules 3 and 4.

3. If the full title (field 245) is present, use the first 100 characters as the title. If not, apply rule 4.

4. If the uniform title (field 240) is present, use the first hundred characters. Furthermore, if the corporate name field (field 110), is present, add it to the end of the uniform title. If the corporate name field is not present but the conference or meeting name (field 111) is, add the conference or meeting

name to the end of the uniform title to form the title. The purpose of adding the main entry is to augment uniform titles consisting solely of such words as "Journal" or "Proceedings".

The LC call number, Dewey call number, title, and LC card number fields include the indicators and subfield delimiters of the original MARC fields. A record of exactly which fields were used to create a given extract record, is not kept.

Parsing

The PARSING program transforms input data into a standard form that is readily sorted and compared by computer. It is used to parse LC call numbers, LC card numbers, prices, and imprint dates in the BPR files, and to parse LC call numbers on the OCLC Distribution tapes.

This program has been enhanced since the previous project to handle more complex parsing algorithms. Specifically, it can now extract the ending value of an LC call number range no matter how the range is specified. This is necessary due to the over nine thousand LC call numbers and ranges that make up the LC/HEGIS transaction.

To produce the ending number in a range of LC call numbers requires "backing up" when a hyphen is encountered to pick up the implied prefix. In QA 43-45 for example, QA is the implied prefix for the ending number, QA 45. The following examples illustrate the variety of forms of LC call number ranges that can be processed using a single parse table.

	LC Call Number Range	Ending Call Number
1.	LB 1028.3	LB 1028.3
2.	Z 7164.E35-.E37	Z 7164.E37>
3.	LC 6571-6581	LC 6581 >
4.	LC 6571-81	LC 81 > (N.B. ambiguous input)
5.	QA 401.2-402.5	QA 402.5 >
6.	LC 158-158	LC 158 >
7.	Q 295	Q 295

It must be remembered that the PARSING program is table-driven so that many other parsing algorithms that require "backing up" could also be implemented using this program (ranges of years, prices, etc.)

The greater than sign (>) at the end of some of the call numbers causes all LC call numbers that match up to the greater than sign to be considered part of the range. For example, LC 158-158 would include LC 158.A-.Z. On the other hand, Q 295 would include only records whose call numbers are precisely that.

Classify

This program maps input records to classes in a hierarchy. In the current project, it was used to assign OCLC and BPR bibliographic records to four digit HEGIS

categories, two digit HEGIS categories, and the ten SUNY groupings of HEGIS based on LC call number.

The CLASSIFY program was written for the previous project but has been substantially changed. The most important change is the ability to define classes in terms of ranges of input values that overlap. The previous version of the program would have required that all LC call numbers and ranges in the LC/HEGIS translation be distinct from one another - that is, not contain the same LC call number. Since many different subject specialists were working simultaneously to specify LC numbers for their specialty areas, such a constraint would have been impossible to maintain. Hence, the program was modified.

Another capability that was added late in the project was the creation of a default class for titles that did not fall into any of the LC call numbers or ranges, thus providing a means for displaying titles that fell through the LC/HEGIS translation.

A third capability was the production of both detail and aggregate records. Detail records contain data pertaining to specific records such as author and title, while aggregate records simply record the count and weight of the input records. The count is the number of input records processed, while the weight may be any numeric quantity, such as price or number of volumes. Detail records may be subsequently sorted to produce lists of bibliographic records by HEGIS category as one example, aggregate records are used in statistical reports.

Aggregate

The purpose of the AGGREGATE program is to accumulate the counts and weights of input records by subclass within class (e.g. by four digit HEGIS within two digit HEGIS), and to produce the following statistics:

1. total count overall - the total number, over all classes, that were assigned of input records that were assigned to any class at all. This does not include titles that fall through the LC/HEGIS translation.

2. other count - the number of input records that are not assigned to any class.

3. total weight overall - the sum of the weights (e.g. prices, number of volumes) of the records counted in 1.

4. other weight - the sum of the weights of input records counted in (2).

5. class/subclass count - the number of input records assigned to a class, or to a subclass within a class.

6. class/subclass weight - the sum of the weights of the input records assigned to a class or subclass.

7. average weight - the weight divided by the count for each pair of statistics above.

8. weight percent of the class - the percent a subclass weight represents of a class' weight.

9. weight percent overall - the percent a class or subclass weight represents out of the overall weight.

10. count percent of the class - the percent a subclass record count is of the class' count.

11. count percent overall - the percent a class' or subclass' count is of the overall count.

The concepts of count and weight enable one to distinguish between, for example, the cost to acquire a group of titles and the number of titles acquired. The percents based on count provide a different perspective from percents based on weight; the former considers each title to be equal. Overall statistics percentages provide a percentage relative to a common base, while class statistics allow one to limit the context to a smaller portion of the collection.

Report

The REPORT program simply generates a formatted report of aggregate statistics or detailed information about titles. If an input field contains a class id, the program looks up the textual description of the class in the class descriptors table of the appropriate hierarchy. Subclass descriptors are indented four positions from class descriptors to differentiate between major groupings and the minor groupings of which they are made up.

This program is the means for maintaining a catalog of hierarchies through additions, deletions, and updates on existing hierarchies.

Since a hierarchy represents a model of the structure being studied, it is desirable to maintain the hierarchy in such a way that the model may be changed easily. Furthermore, it may be desirable to keep several variations of the same hierarchy to accommodate similar yet distinct models. Hence, the idea of maintaining a permanent record of hierarchies was formed. This "catalog" of hierarchies may be perused for the hierarchy that is closest to the structure of the model being studied. It may then be modified as necessary to create a new hierarchy which is added to the catalog. This encourages both uniformity and diversity at the same time.

The LC/NEGIS translation along with the two digit NEGIS and ten SUNY groupings of NEGIS make up the primary hierarchy in the current catalog. Other modifications to it may easily be accommodated by adding new hierarchies to the catalog.

Combine

The COMBINE program matches up to ten input files on the same keys and creates output records consisting of selected fields from each file.

The program was designed primarily initially to perform the matching of OCLC and BPR records on LC card number. Due to the general usefulness of the program, its design was broadened to allow matching of any files so long as they contain fixed length records. The program has been used not only to match OCLC and BPR records but also to combine the reports that resulted from the separate analysis of the various sources of data.

BPR Extract

This program was written to extract data from the file of Book Publishing Record data provided to the project by R. R. Bowker Company. The program extracts fields according to their "tag" much like the EXTRACT program does for OCLC tapes.

Some fields occur more than once in the same title (e.g. author). The program handles multiple occurrences of the same field in one of two ways at the option of the user. The option used in this project's analyses selected only the first occurrence of each field. The alternative is to "explode" the record, creating one output record for each combination of multiply occurring fields.

Select A

This program selects OCLC records on the basis of their sequence number on the distribution tapes. The EXTRACT and INITIAL EXTRACT, and BPR EXTRACT programs all include the sequence number of the OCLC record from which data is being extracted as part of the output record. In combination, these programs provide a feedback mechanism for obtaining additional data from OCLC records. That is, suppose for example that only the imprint date is extracted from each OCLC record initially. Upon analysis, one may wish to see the authors and titles of records for a specific year. The SELECT program could be used to create a new file of OCLC records published in the desired year, and the EXTRACT program could then be used to select the necessary fields from the new file. Since the cost of using the more flexible EXTRACT program is proportional to the number of fields extracted, this also provides a way to reduce the cost of analysis.

Select B

There are actually two versions of the SELECT program. The older version uses the type of record sequence numbers that are generated by the EXTRACT program, while the newer version uses record sequence numbers as generated by the INITIAL EXTRACT program which are somewhat more efficient to work with. In addition, the newer version of SELECT may be used on non-MARC formatted files such as the BPR files. The only restriction is the records may not exceed 2,048 characters in length.

Contains

This program generates the contains relationships table for a hierarchy. The contains relations table is a list of the classes that contain - that is, are above - another class in the hierarchy. For example, in a REGIS hierarchy, each four digit REGIS category would be contained by a two digit REGIS category. The purpose of creating the contains relations table is to expedite the process of assigning titles to components of the collection.

Print Wave

This program lists the contents of an OCLC record. It is essentially the same as an earlier program written by Philip L. Long, who was at one time in the SUNY Central Administration Office for Library Services.

Partition

This is a program that was not used in the current project because its functions were subsumed by the CLASSIFY program. It was the original program for assigning titles to components of the collection based upon input value (e.g. LC call number). Its primary drawback is that it does not allow the ranges of input values, which control the mapping of titles, to overlap with one another. For some applications, however, such as a simple division of the collection by LC classification schedules, this program may be more efficient.

Rank

This program also was not used in the current project. It ranks classes components of the collection according to a numeric measure. Classes with the same numerical weight may either be assigned the same rank, or different ranks at the option of the user.

Other Programs

The remaining programs are not of general utility and therefore are not described in detail. Their uses are described as follows:

1. print selected pages of a report.

2. add a holding library code to OCLC/MARC records without a holding library symbol in the holdings (049) field.

3. create microfiche from a tape containing a report.

4. check LC call number ranges to ensure that the ending value is not less than the starting value.

5. compare files record by record to check for errors in file copying.

6. list the last few records of a file in reverse order.

7. perform updating of source programs by inserting and deleting lines of text.

6. REPORTS AND ANALYSES

Data Selection

The data used in the reports presented in this chapter have three sources:

1. A selection from the SUNY/OCLC Combined Distribution Tape representing the OCLC cataloging of 9 SUNY Colleges of Arts and Sciences, 2 SUNY University Centers, 3 Specialized Colleges, and 1 Agricultural and Technical College between July, 1977 and July, 1978.

2. SUNY/CASA Institutional Data from the 15 SUNY campuses represented on the SUNY/OCLC Combined Distribution Tape. This file includes courses offered, enrollments, course level, student level, faculty workload, faculty full time equivalents, student credit hours generated, etc.

3. Book Publishing Record (BPR) files which are the property of the R. R. Bowker Co. These files contain 101,410 bibliographic records as published in the 1975, 1976, and 1977 editions of BPR.

Data selection criteria are described below.

SUNY/OCLC Combined Distribution Tape.

The entire tape, representing cataloging on OCLC during July, 1977 through July, 1978 for all libraries subscribing to the OCLC Combined Tape, contains 781,247 bibliographic records.

The portions extracted from the OCLC record were:

1. OCLC Record Number

2. Holding Library Code

3. Characters 5 - 7 of the leader:

   Record Status
   Type of Record
   Bibliographic Level

4. Characters 17 - 18 of the leader:

   Encoding Level
   Descriptive Cataloging Form

5. Leader 22

   Transaction type (produce, update, replace, cancel, all produce)

6. 008 Field

7. Record sequence number (this number identifies the OCLC record from which the data are extracted)

8. LC call number, first fifty characters

9. Dewey call number, first fifty characters

10. Title, first hundred characters

11. LC card number

12. LC parsed call number (generated by SUNY programs)

All records indicating a transaction type of "cancel" or "update" were eliminated. All records for libraries other than the 15 SUNY libraries described above were eliminated leaving a total of 185,832 records. When a single OCLC record was used more than once by a holding library, only the latest use was incorporated into the data file. The number of unique OCLC data base records used was 139,215. The difference between 185,832 and 139,215 reflects the use of the same OCLC record by more than one institution.

Quality of data. The quality of the OCLC data reflects the quality of cataloging in the individual libraries. No attempt was made to establish special criteria or standards for cataloging input between July, 1977 and July, 1978. The libraries were not told that their cataloging data would be used for such a purpose, but libraries are aware that they are creating an archive of all cataloging for future applications such as collection analysis, COM catalogs, on-line catalogs, etc. For this reason and others, which will be discussed, the libraries shall only be identified by alphabetic codes.

Book Publishing Record Data.

The R. R. Bowker Company generously supplied a tape containing certain data from its 1975, 1976 and 1977 Book Publishing Record files. The following fields were extracted from the BPR tape:

1. LC class number
2. Dewey class number
3. LC card number
4. Format Field (contains codes for fiction/non-fiction, type of main entry, intellectual level, source of cataloging, etc.)
5. Imprint date
6. Price
7. Title
8. Subject heading (Primary)
9. Tape sequence number

An SPSS analysis was run on some characteristics of the BPR data.

Imprint date in relation to BPR date. A crosstabulation was done to establish the imprint dates of titles contained in the three volumes of BPR. Since the project staff was interested in estimating the current publishing output, the imprint date rather than the BPR date was used as a selection criterion. Within any given BPR year, approximately 75% of the records have an imprint date identical to the BPR year. Approximately, 25% of each year's publications are reported in BPR the year subsequent to their imprint. Books published prior to 1975 were eliminated from the analyses. There were a total of 9677 titles published prior to 1975 in the data base, 9150 of which were 1974 imprints. There were also sixty 1978 imprints which the project staff eliminated. Only titles with an imprint date of 1975-1977 were used in average cost computations.

Intellectual level and "nature" of BPR data. The BPR data file contains a code indicating juvenile or general level. Juvenile level materials represent 7.2% of the

titles in the 1975 through 1977 BPR data base. It was not possible to distinguish between juvenile fiction and non-fiction. Since the number of juvenile titles was relatively small, and may be acquired by some academic libraries it was decided to retain these in the data.

The BPR data base also identifies five categories labeled as the "nature" of a title. The distribution according to type of material follows:

		Percent	Titles
1.	Fiction	13.3 %	13,500
2.	Non-fiction	79.7 %	80,850
3.	Auto-biography	2.1 %	2,126
4.	Biography	3.6 %	3,677
5.	Collective biography	1.2 %	1,257

It is important to be aware of these characteristics in interpreting the reports and to allow for material which may be of a general rather than scholarly nature. It is not possible to distinguish between popular and scholarly publications in either the BPR or OCLC data base. The discrepancy between titles published and titles appropriate for academic library purchase may be discerned as OCLC records and records of academic library jobbers are compared with BPR and other publishing records, but testing for this was not addressed in this project.

Source of cataloging data in BPR. MARC tapes distributed by the Library of Congress are the primary source of BPR cataloging records. The remaining records represent original cataloging by Bowker.

Distribution of Cataloging Sources

	Titles	Percent
MARC	88928	87.8 %
Original	12321	12.1 %

The BPR data represent the majority of U.S. commercial publishing output for the years 1975 through 1977. Report - 13 (projected) relates the BPR data to the titles 15 SUNY libraries are cataloging. An analysis of the titles acquired by the library but not found on the BPR file (Report 14 projected) will help to identify other sources of publishing data which ought to be considered in estimating the body of materials from which libraries are selecting.

Price of BPR titles. An SPSS breakdown of price records of all BPR titles for each BPR year revealed the following information:

Year	Sum	Mean Price	Std. Deviation
1975	404,349	$ 14.31	30.36
1976	410,466	$ 14.25	31.12
1977	488,565	$ 15.02	39.87
Total	1,303,380	$ 14.55	34.37

Records with no recorded prices: 11841 or 11.7% (These records were omitted in analyses involving prices)

The staff sought to eliminate records with prices which were out of line with the norm. It was determined that records with prices higher than $95.00, or is approximately two standard deviations above the mean price, should be excluded when calculating average prices in Reports 11 and 12.

Paperback prices. The BPR data files include both hardcover and paperback editions. Initially, the exclusion of all paperbacks was contemplated. However, academic libraries do purchase many paperbacks: 1) due to unavailability of hard cover editions and 2) due to cost savings on paperback editions. Therefore, it was deemed inadvisable to eliminate all paperback editions. In BPR records encompassing multiple editions of the same work, the first price given is for the hardcover edition. Therefore, by selecting the first price, the price is the hardcover edition price except in cases where there is no hardcover edition. This procedure partially accounts for an average price different from that reported in the annual book publishing statistics which record hardcover and paperback editions separately.

Titles versus volume price. The annual book publishing statistics report cost per volume. However, the cost field in the BPR records is somewhat difficult to interpret. In some cases, the price is reported as x dollars per set, sometimes as x dollars per volume. It was the intention of the project staff to report all publishing data by title since title represents the intellectual unit of information. Volume is considered a packaging device. Cost per title was seen to be more relevant to library collection development than cost per volume. Although Bowker reports average price by volume, the procedure for doing this by HEGIS category could not be duplicated for the project. Therefore, the price given in the BPR record is accepted as is. There is a downward distortion in prices when the price is listed "per volume" because fiction and juvenile literature are included. In the case of large, expensive sets, the price will probably exceed the $95.00 cut-off point established and will be eliminated from the calculation of average prices. Budget planners should be aware of these distortions in the data when estimating average book prices by subject.

Records with missing price data. Approximately 11% of the BPR records contained no price data. These included several categories:

1. Gratis
2. Price not set
3. Price not available

Although Bowker continues to add price information to these incomplete records after the BPR volume is published, the data tapes available for this project were the tapes used to print BPR and contained no updates subsequent to publication. The relatively high number of records lacking price data presents a problem in establishing average prices of publications for the three years surveyed. Further work, analyzing these records to determine the proportion of gratis titles included in these "price missing" records as well as the prices which were eventually recorded or set and their effect on the average price in each subject, is desired.

The "price missing" records are excluded in reports of average price per subject, but are included in the matching of OCLC titles against BPR titles. (Report 13) However, when the average cost of matched titles is calculated (Report 7), the "price

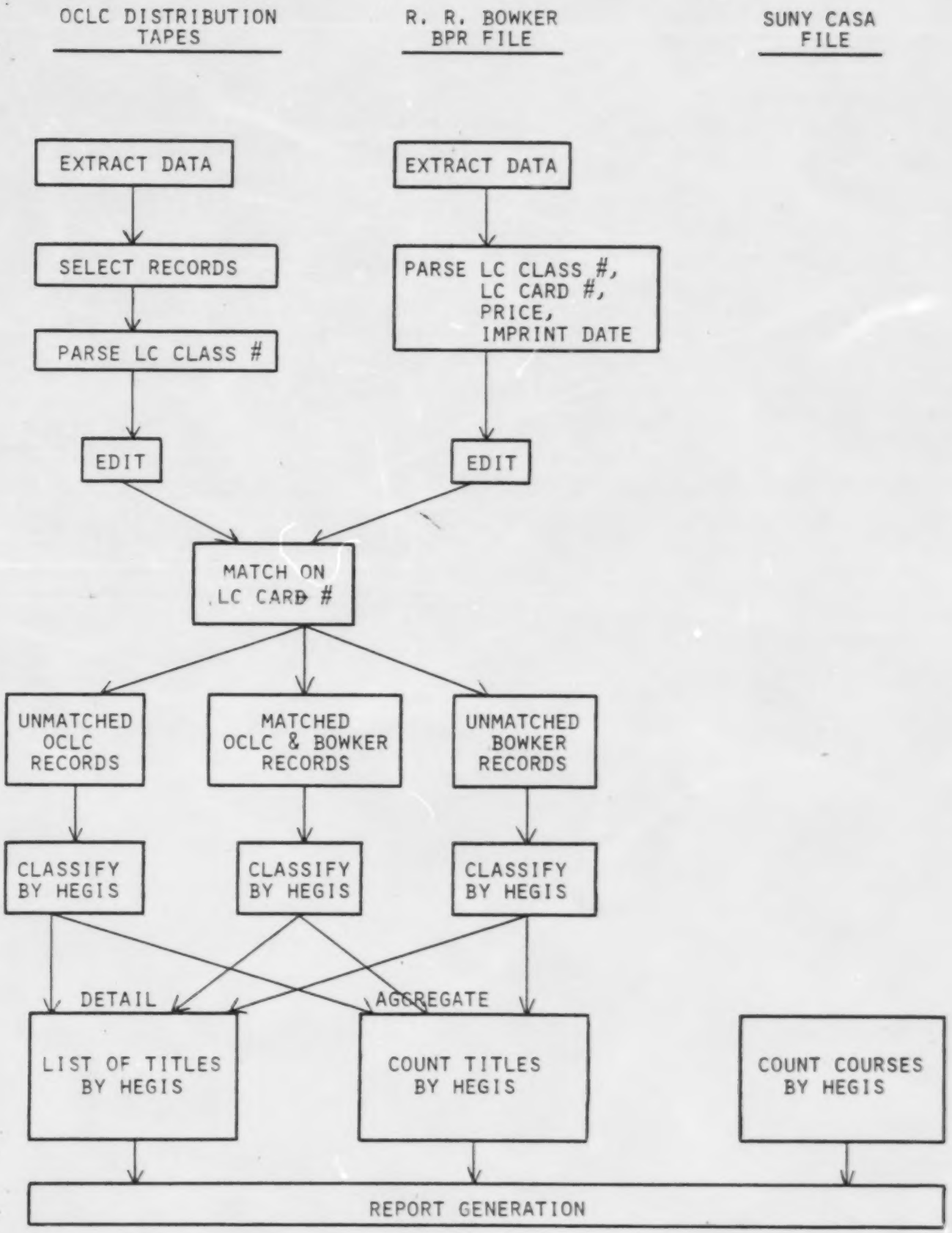
missing" titles are again excluded.

General considerations.

Project staff outlined a number of possible reports designed to demonstrate the capabilities of the programs, utilize the LC/HEGIS Translation, and assist in its evaluation. The reports were designed to anticipate the reports collection development personnel at individual libraries and at system-wide administrations would find useful. They were, by no means, designed to exhaust the potential of the system or to be used, at this early stage, to make policy or financial decisions. One of the most productive aspects of the MIS is the ability to generate successive reports or to "chain" a series of reports. One may want to do this to explain some feature appearing on a report, e.g. matching OCLC cataloging records against BPR publishing data and then generating a second report on the cataloged titles which were not matched in BPR to determine whether they were outside the scope of BPR such as foreign publications or older titles. This chaining process, and indeed, the entire process of report generation is relatively simple, since once the data elements are extracted and the variables are defined, most report generation can be done using the Statistical Package for the Social Sciences (SPSS) which can be used by individuals having little programming background.

Figure 5-1 demonstrates the steps involved in transforming the three strands of data described in the introduction to this chapter into the various report formats.

FIGURE 5-1


OCLC DISTRIBUTION TAPES
R. R. BOWKER BPR FILE
SUNY CASA FILE
EXTRACT DATA
EXTRACT DATA
SELECT RECORDS
PARSE LC CLASS #, LC CARD #, PRICE, IMPRINT DATE
PARSE LC CLASS #
EDIT
EDIT
MATCH ON LC CARD #
UNMATCHED OCLC RECORDS
MATCHED OCLC & BOWKER RECORDS
UNMATCHED BOWKER RECORDS
CLASSIFY BY HEGIS
CLASSIFY BY HEGIS
CLASSIFY BY HEGIS
DETAIL
AGGREGATE
LIST OF TITLES BY HEGIS
COUNT TITLES BY HEGIS
COUNT COURSES BY HEGIS
REPORT GENERATION

OCLC Data. After extracting data from the OCLC tapes and selecting the records of the fifteen SUNY campuses, the LC class number was parsed into a standard format that could be used for the translation from LC to HEGIS.

The LC card number was already in a standard form so it did not require parsing. The standard form consists of two digits for the year, followed by six digits for a sequence number within year.

Having parsed the LC class number, the extracted OCLC data was then sorted by LC card number for matching with the BPR data. OCLC records having no LC card numbers were separated from the file being sorted.

BPR Data. After extraction, the LC call number, LC card number, price, and imprint date of the BPR data were parsed, and the BPR year was added to each record. The combined files for 1975-77 were then sorted on LC card number. Those records without an LC card number were separated into another file before sorting.

CASA Data. The CASA file contains data on every course taught in the SUNY system. A HEGIS code is assigned to each course that reflects the content of the course rather than the department offering it.

The CASA file was aggregated by HEGIS category within campus to give the following statistics for lower division, upper division, and graduate courses:

1. number of courses
2. student credit hours
3. faculty contact hours
4. enrollment
5. FTE faculty.

These factors were selected with the intent to do a regression analysis using them as the independent factors, and the number of titles acquired as the dependent variable. At the suggestion of the Advisory Group, total salary by HEGIS category was also added. Since the regression was not completed, we do not know whether these factors are appropriate for predicting patterns of library acquisitions.

Matching OCLC and BPR Data. OCLC and BPR data were then matched on LC card number. This resulted in three files of data - matched records, unmatched records, and unmatched BPR records.

From the fundamental building blocks created in this project, it is possible to discern and define an almost infinite number of reports. Fifteen were defined in a first group to be developed, and programs written in modular form through which the reports could be generated. In the event, however, because of the pressures of time it was only possible actually to run a few of the projects. The remainder of the chapter is devoted to describing the preferred reports, and each specific report is annotated as to whether it was run, and whether the data shown are actual or examplar.

Summary of Reports included in Chapter 6

Group A.	Performed during project
1/2	Titles acquired in relation with Student Credit Hours Generated
3	List of Ranges and Number of titles caught by each
4	List of titles by Hegis, translation table from campus OCLC tape.
5	List of BPR titles by Hegis translation table.
6	Overlap analysis of titles reported in BPR.
7	Average Price of BPR titles by Hegis Subject.
Group B.	Not performed during project, but similar to Reports in Group A.
8/9	Titles published vs Student Credit Hours Generated (See Reports 1/2)
10	List of Ranges and Numbers of BPR titles caught by each.
Group C.	Not performed during project, illustrated with proposed formats and hypothetical data
11	Overlap among 4 digit HEGIS categories in OCLC Records (proposed)
12	Disciplinary Price Index (proposed)
13	Matching titles cataloged by libraries with titles reported in BPR, analyzed by Hegis code (proposed)
14	Analysis of Library Titles which do not match BPR records (proposed)
15	Regression analysis of relationship between selected academic factors and library titles cataloged (proposed)

Report 1/2. Titles Acquired in Relation to Student Credit Hours Generated.

Data. These reports record the number of bibliographic records entered on the SUNY/OCLC Combined Distribution Tape during the period July 1, 1977 through July 1, 1978 and the numbers of student credit hours generated recorded on the SUNY Course and Section Analysis (CASA) of Fall, 1977. The data represent 9 of the 12 SUNY Colleges of Arts and Sciences, two of the four SUNY University Centers, three specialized colleges and one Agricultural and Technical College.

Purpose. The titles acquired and academic data are analyzed for each campus according to three sets of subjects, i.e. four digit HEGIS categories, two digit HEGIS categories and by a ten discipline grouping consisting of four digit HEGIS categories and designed by SUNY for budget purposes. Finally, all of the instructional subjects are aggregated to produce a grand total for each campus.

The academic data in the CASA file are permanently linked to HEGIS categories and can easily be aggregated. The bibliographic data are sorted into HEGIS categories and their aggregations by means of the LC/HEGIS Translation. The LC Class number of each record is sorted and matched against the translation tables and sorted into each instructional subject. One should note that the grand totals are smaller than the sum of the individual four digit HEGIS, two digit HEGIS and 10 Disciplinary Groupings. The counts for a two digit HEGIS category are less than the sum of the individual four digit HEGIS categories which it contains. The reason for this seeming inconsistency is that a single record may have a classificaiton number which is mapped to more than one HEGIS category or Instructional Subject. As aggregation occurs, the record is counted only once at each successive level of aggregation so that a title which may have been counted in both Zoology and in Entomology at the four digit HEGIS level is counted only once when Zoology and Entomology are aggregated into the two digit level Biological Sciences.

Factors influencing interpretation. Note must be made that the translation is incomplete in some subjects: notably area studies and health professions. This means that no titles are reported for those subjects. At the four digit HEGIS level, the omitted subjects are reported as zero titles. At the two-digit and then discipline levels, however, the titles are only those which map to categories in the aggregation represented in the translation. For instance, due to logistical problems in Area Studies 03, Latin American Studies and African Studies are the only area studies which have been included in the present translation; however at the two digit HEGIS level all area studies student credit hours are counted. Therefore, there is little correspondence between individual subjects reported from the CASA file and the subjects reported from the OCLC file. Refer to Appendix 1 for a complete report on which subjects are included in the LC/HEGIS Translation.

It should be pointed out that neither the classification of courses by HEGIS number nor the LC/HEGIS Translation have been adjusted to assure that the Translation fully represents the subjects taught at a given campus. This problem is reported in the Data chapter. In this report the classification of courses and the mapping of LC to HEGIS are assumed accurately to match bibliographic records with student credit hours.

In the campus specific report, no four digit HEGIS categories which are not offered in the fall semester of 1977 are printed. Therefore all of the titles which fall into ranges which are mapped to HEGIS categories not reported by the campus appear at the end of the report in the "other" category. It is anticipated that as institutional data is used by librarians in library management information systems, communication

between librarians and institutional planners will increase. Institutional researchers will gain improved feedback on the results of course and program classification, and librarians will learn more about the curricular offerings of their campuses.

Analysis at two digit HEGIS and the 10 SUNY aggregate discipline levels may show greater correspondence between programs and acquisitions due to greater flexibility within categories. Given the inconsistency in level of detail noted in the earlier discussion of HEGIS, the reports may be more useful if four digit HEGIS categories were aggregated into an entirely different set of instructional subjects. It may be advantageous to propose that a set of four digit subjects be constructed to attain consistent levels of detail, i.e. each one could encompass an equivalent portion of knowledge and each instructional subjects could resemble a traditional listing of academic subjects. One virtue of the computer programs is that such a table can easily be compiled and maintained. The process can be repeated and evaluated until the best subject grouping for any campus is achieved.

In the analysis of titles the "other" category represents those records on the distribution tape whose classification numbers have not been mapped to any HEGIS category. Since the translation has not been extensively tested and evaluated, ranges may be omitted which should have been mapped to HEGIS. Continued development will improve on this imperfect situation.

Interpretation of Results of Reports 1 and 2 and possible applications.

These reports are presented for descriptive purposes. Previous discussion of the relationship between campus academic programs and library acquisitions patterns indicates that decision makers maybe interested in seeing these two factors presented side by side.

It is not implied in this report that any causal or linear relationship exists between these two factors. Neither should one imply that extensive purchases in subjects which apparently do not correspond with subjects taught at the campus are to be criticized. The possible mis-matches between campus reporting and library collections via the LC/HEGIS translation have already been discussed. It will not be possible to examine potential linkages between cause and effect until data from a variety of campuses over different periods of time have been processed, evaluated and reviewed.

COMPARISON OF TITLES CATALOGED WITH STUDENT CREDIT HOURS
IN ONE SUNY INSTITUTION

HEGIS Group	Number of Titles	Percent of Titles	Number of Student Credit Hours	Percent of Student Credit Hours
Education				
08 Education	455	9.60	2,065	3.1
Math and Computer Science	369	7.80	13,179	19.8
07 Computer & Info. Sci.	112	2.30	3,563	5.3
17 Mathematics	345	7.30	9,616	14.4
Physical Sciences and Engineering	811	17.20	11,614	17.4
02 Architecture & Envirn. Sci.	161	3.40		
09 Engineering	376	8.00		
19 Physical Sciences	432	9.10		
Psychology				
20 Psychology	1646	35.00	7,569	11.3
Social Sciences	3066	65.20	31,994	48.1
Total	4,696	99.9	66.421	100.

Data: SUNY OCLC Distribution Tape, Jan.-Mar. 1977. Sorted by Partial LC/HEGIS Translation, May, 1978. All subjects are not represented in this analysis. Student credit hour data from SUNY Course and Section Analysis file, Fall, 1976.

EXAMPLE REPORT 1/2

Report 3. List of Ranges with Number of Titles Caught by Each

Data. The data used for this report are the combined SUNY cataloging records described in the introduction to this chapter.

Description. Each of the individual ranges of the LC/REGIS Translation are listed along with the number of titles identified by that range. Due to the way the Translation is constructed (See LC/REGIS Translation chapter) the same records may fall into a number of ranges; thus the total number of titles caught will reflect a degree of overlap.

Range size. There is great disparity in the size of the ranges; some ranges include several hundred numbers, whereas other ranges consist of a single decimal number. It is difficult to draw comparisons between such disparate ranges. Three range types were identified to assist in grouping the ranges by relative size. These three types are separated on report 3 as whereas Report 3b displays the data on all ranges in descending order of frequency. In the future, it may be possible to provide an more satisfactory measure of the size of each range in order to compare the titles caught by size of range.

Applications. The primary purpose of this report is to assit in the evaluation of the LC/REGIS Translation. It indicates which ranges are not catching any titles. This may indicate that the subject area is not currently productive, that none of the libraries are buying material in that area or simply that the range is so small that the likelihood of titles being caught is miniscule. Another possibility is that other portions of the LC Schedule overlap with this range and are preferred by catalogers. One might infer that ranges which are not catching any titles could be eliminated from the translation without affecting the results. This action would eliminate the possibility of identifying unproductive or neglected subjects.

Future work. One of the initial reasons for identifying rather specific ranges and providing descriptors was to enable libraries to analyze their collections in detail. In the subject of law, for example, REGIS provides only one category. The translation provides over 300 individual ranges relating to law divided by country and then by legal topic. These ranges were designed to relate to the typical law curriculum and topics as they are conceived by legal experts. There are many cases in which the LC classification separates topics which legal experts would combine e.g. education law in LB and education law of the U.S. in KF. The original plan was to group the ranges into logical categories within the four digit REGIS groupings. Time has not permitted this grouping this year, but the potential still exists.

Another subject composed of a multitude of minor ranges is Ethnology. The ethnology of individual countries and states is found as a sub-division of the history and culture of countries in the D, E, and F classifications of LC. It is impossible to combine all the numbers for ethnology into a single range. Each of these numbers retrieve very few titles individually, but together they represent a sizeable group of titles on the ethnology of individual countries and states.

EXAMPLE REPORT 3

490100	A		
490100	AM	0001	
490100	AN		
150900	AT	0108	
490300	AZ		
490100	AS		
150900	B	0020	
150900	P	0031	
150900	B	0053	
150900	B	0068	
150900	B	0069	
150900	P	0171	
150930	B	0165	
150900	B	0504	
150900	B	0670	
150900	P	0108	
150900	B	0720	
150900	P	0770	
150900	P	0870	
150900	B	0823	
030800	B	1001	
150900	P	0256	
150900	PC	0050	
150900	PC	0060	
150900	PC	0060	
150900	PL	0101	
150900	PC	0135	
150900	PC	0135	
150900	PC	0135	
150900	PC	0135	
040100	PC	0177	
090900	PC	0177	
150900	PC	0177	
150900	PD	0010	
150900	PD	0041	
150900	PD	0045	
150900	PD	0131	
150900	PD	0100	
150900	PD	0125	
150900	PD	0125	
150900	PD	0150	
150900	PD	0240	
150900	PD	0444	
150900	PD	0444	
150900	PD	0445	
150900	PD	0450	
150900	PD	0451	
150900	PF	0334	
150900	PF	0334	
150900	PF	0014	
150900	PF	0014	
150900	PF	0018. 5	
150900	PF	0028. 5	
150900	PF	0028. 8	
150900	PF	0032	
150900	PF	0051	
150900	PF	0171	
150900	PF	0175	
150900	PF	0176	

AI		
AM	0151	
AP		
AT	0108	
AZ		
AI		
B	0029	
B	0048	
B	0067	
B	0068	
B	0105	
B	0162	
B	0491	
B	0670	
B	0708	
B	0308	
B	0765	
B	0785	
B	0870	
B	0823	
B	1004	
B	5739	
PL	0051	
PL	0070	
PL	0099	
PL	0117	
PL	0135	
PL	0135	
PL	0135	
PL	0135	
PL	0177	
PL	0177	
PL	0199	
PD	0020	
PD	0041	
PD	0110	
PD	0151	
PD	0151	
PD	0175	
PD	0175	
PD	0236	
PD	0241	
PD	0241	
PD	0445	
PD	0445	
PD	0445	
PD	0450	
PD	0455	
PF	0039. 5	
PF	0039	
PF	0039	
PF	0039	
PF	0039. 5	
PF	0078. 5	
PF	0078. 8	
PF	0080	
PF	0171	
PF	0175	
PF	0175	
PF	0176	

NUMBER TITLES
46
4
7
1
16
24
14
7
32
2
30
32
31
15
2
100
16
0
1
1
2
430
1
4
2
3
0
0
0
0
0
0
0
30
12
1
4
1
3
13
13
59
5
5
17
11
11
53
10
20
20
20
20
32
14
6
3
144
10
10
6

Report 4. List of Titles by HEGIS (Instructional Subject) Using SUNY OCLC Data.

(a) Format 1

(b) Format 2

Description and Purpose. Report 4 is an expansion of Reports 1 and 2; it lists the library holdings which are merely counted in earlier reports. A list of this sort would not be routinely requested by a library. Produced for specific purposes, however, it could be extremely useful. It might be used to support an accreditation review or new program justification by providing a subject listing of library holdings.

Two separate formats are illustrated.

Format 1 reports (in column order) LC Class Number/Cutter/Title /first 90 characters, including delimiters B and C/ first imprint date, each line is ascending Call # order within Hegis 4 digit Group.

Format 2 reports Class Number/HEGIS Code/First and Second Indication/Title /First 86 characters of title, including delimiters 'a', 'b', and 'c/ OCLC #. These test data were taken from a multi institutional tape from the OCLC system. Multiple appearances of the same record indicate holdings of the same title in more than one library. The libraries are not identified in this report.

The immediate purpose of this report and the corresponding listing of BPR titles (Report 5) is to evaluate the LC/HEGIS Translation. This report will be studied carefully to identify problems in the translation. These would be of two types: 1) titles which fall into HEGIS categories which they should not fall into. 2) titles which fail to fall into HEGIS categories they should fall into. There are at least two reasons for each occurrance: 1) the title has been incorrectly classified or 2) the range has been incorrectly mapped to a HEGIS category.

Extra attention should be paid to the "others" category, since these are titles which are apparently supporting no instructional subject. These will be carefully screened to determine whether some new ranges ought to be considered for the translation. This run will also reveal new LC numbers which were not available when the subject specifalists did their coding.

A secondary, but potentially important use of Report 4(a) and (b) and Report 5 is the comparison of LC class numbers and LC Subject Headings. A list of this sort could be a useful reference tool to catalogers and an evaluation tool for the Library of Congress to determine the consistency with which call numbers and primary subject headings are being applied. The subject heading reported is the primary subject heading.

EXAMPLE REPORT 4A (FORMAT 1)

10/17/78 LIBRARY HOLDINGS BY DISCIPLINE FOR THE COLLEGE AT UTICA /ROME — PAGE 124
CATALOGED SINCE LAST JULY 1, 1977

DISCIPLINE BUSINESS AND MANAGEMENT

MEJIS CATEGORY 0504 BANKING & FINANCE

LC CLASS NUMBER	CUTTER	TITLE	IMPRINT DATE
HA 0539	.E941954	THE THEORY OF INTEREST AS DETERMINED BY IMPATIENCE TO SPEND INCOME AND OPPORTUNITY TO INVEST	1954
HA 0539	.H48	THE TERM STRUCTURE OF INTEREST RATES; BFINANCIAL INTERMEDIARIES AND DEBT MANAGEMENT C BY JA	1971
HC 0079.C3	M37	THE MEASUREMENT OF CAPITAL ; BTHE METHODOLOGY OF CAPITAL STOCK ESTIMATES IN OECD COUNTRIES /	1976
HD 0052	.B51971	THE CAPITAL BUDGETING DECISION; BECONOMIC ANALYSIS AND FINANCING OF INVESTMENT PROJECTS C BY	1971
HD 0052.5	.B71973	THE INDUSTRIAL BUYING DECISION; IMPLICATIONS FOR THE SALES APPROACH IN INDUSTRIAL MARKETING	1973
HD 2741	.A285	ACQUISITION BEHAVIOR OF U.S. MANUFACTURING FIRMS, 1946-1965 C BY H. IGOR ANSOFF AND OTHERS	1971
HD 2741	.B33	BUY, SELL, MERGE; HOW TO DO IT C BY WILLIAM H. BAUMER AND LEO J. NORTHART.	1971
HD 2741	.B61959	ORGANIZING AND FINANCING BUSINESS C BY JOSEPH H. BONNEVILLE, LLOYD E. DEWEY AND MARRY M.	1959
HD 2741	.B851970	EFFECTS OF TAXATION; CORPORATE MERGERS C BY J. KEITH BUTTERS, JOHN LINTNER AND WILLIAM L.	1970
HD 2741	.C744	CONGLOMERATE MERGERS AND ACQUISITIONS; BOPINION AND ANALYSIS.	1970
HD 2741	.H859	INCORPORATING A PROPRIETORSHIP OR PARTNERSHIP.	1966
HD 2741	.M2	MANAGEMENT PROBLEMS OF CORPORATE ACQUISITIONS C BY MYLES L. MACE AND GEORGE G. MONTGOMERY	1962
HD 2741	.M57	TECHNIQUES IN CORPORATE REORGANIZATION; BSELLING AND DIVESTING, SPIN-OFFS AND SHELLS, MULTIP	1972
HD 2741	.S35	ACQUISITIONS, MERGERS, SALES AND TAKEOVERS; BA HANDBOOK WITH FORMS C BY CHARLES A. SCHARE.	1971
HD 2741	.W9	BUSINESS COMBINATIONS; PLANNING AND ACTION C BY ARTHUR R. WYATT AND DONALD E. KIESO.	1969
HD 2965	.A24	COOPERATIVE BUSINESS ENTERPRISE / MARTIN A. ABRAHAMSEN.	1976
HD 3446	.A61976	COOPERATIVES ; BDEVELOPMENT, PRINCIPLES, AND MANAGEMENT / CEWELL PAUL ROY.	1976
HF 5550	.C49	ZERO-BASE BUDGETING COMES OF AGE ; BWHAT IT IS AND WHAT IT TAKES TO MAKE IT WORK / CLOGAN M.	1977
HF 5550	.W19	FINANCIAL MANAGEMENT OF THE SMALL FIRM / CERNEST W. WALKER, J. WILLIAM PETTY, II.	1976
HF 5550	.W443197C	BUDGETING ; BPROFIT PLANNING AND CONTROL / CGLENN A. WELSCH.	1976
HG 0065	.G62	FINANCIAL MARKET PLACES; BA DIRECTORY OF MAJOR CORPORATIONS, INSTITUTIONS, SERVICES, AND PUBL	1972
HG 0151	.T46	THE THORNDIKE ENCYCLOPEDIA OF BANKING AND FINANCIAL TABLES / CBY DAVID THORNDIKE.	1973
HG 0551	.C631971	MONEY, BANKING, AND THE ECONOMY C BY JOHN A. COCHRAN.	1971

52

57

58

		1509	10 ANTSOLM OF THE WEST: BA HISTORICAL SURVEY OF WESTERN PHILOSOPHY IN ITS SOCIAL AND POLI	03CC345
		1509	11 AIHE GREAT PHILOSOPHERS.	0713061
		1509	14 AIHE ORIGINS OF EUROPEAN THOUGHT ABOUT THE BODY, THE MIND, THE SOUL, THE WORLD, TIME,	6722497
		1509	14 AIHE STORY OF PHILOSOPHY: BIHE LIVES AND OPINIONS OF THE GREATER PHILOSOPHERS, CBY WILL	0004001
		1509	14 AIHE WESTERN PHILOSOPHERS: BAN INTRODUCTION C BY E. W. F. TOMLIN.	0004002
		1509	11 AIHE GREAT PHILOSOPHERS C BY E.W.F. TOMLIN.	0004002
		1509	11 AA HISTORY OF PHILOSOPHICAL SYSTEMS.	1012050
		1509	14 AIHE STORY OF PHILOSOPHY : BIHE LIVES AND OPINIONS OF THE GREATER PHILOSOPHERS / CBY WI	1400004
		1509	14 AIHE PURSUIT OF WISDOM : BA HISTORY OF PHILOSOPHY / CBY JOSEPH MANSFIELD.	4172764
		1509	14 AIHE GREAT PHILOSOPHERS. A SOUND RECORDING	3153596
		1509	12 AA HISTORY OF PHILOSOPHY. CBY B.A.G. FULLER.	3104412
		1509	10 AGLOBAL HISTORY OF PHILOSOPHY / CJOHN C. PLOTT WITH JAMES MICHAEL DOLIN, RUSSELL E. MAI	3342416
		1509	14 AIHE STORY OF PHILOSOPHY : BIHE LIVES AND OPINIONS OF THE GREATER PHILOSOPHERS / CBY WI	3441130
		1509	14 AIHE STORY OF PHILOSOPHY: BIHE LIVES AND OPINIONS OF THE GREATER PHILOSOPHERS / CBY WI	3611141
		1509	04 AIHE PHILOSOPHER'S INDEX : BA RETROSPECTIVE INDEX TO U. S. PUBLICATIONS FROM 1940 / C E	3611161
	B 0074	1509	12 AA STUDENT'S HISTORY OF PHILOSOPHY, CBY ARTHUR KENYON RUGERS.	3033415
		1509	10 ALIVING OPTIONS IN WORLD PHILOSOPHY / CJOHN A. HUTCHISON.	1623074
		1509	10 ALIVING OPTIONS IN WORLD PHILOSOPHY / CJOHN A. HUTCHISON.	2525016
		1509	10 ALIVING OPTIONS IN WORLD PHILOSOPHY / CJOHN A. HUTCHISON.	2525016
		1509	10 ALIVING OPTIONS IN WORLD PHILOSOPHY / CJOHN A. HUTCHISON.	2525016
		1509	10 ALIVING OPTIONS IN WORLD PHILOSOPHY / CJOHN A. HUTCHISON.	2525016
		1509	10 ALIVING OPTIONS IN WORLD PHILOSOPHY / CJOHN A. HUTCHISON.	2525016
		1509	12 AA HISTORY OF PHILOSOPHY / CCLEMENT C. J. WEBB.	2525016
	B 0075	1509	10 AOUTLINE-HISTORY OF PHILOSOPHY C BY WILLIAM S. SAHAKIAN.	3033993
	B 0077	1509	11 AHISTORICAL STUDIES IN PHILOSOPHY. CAUTHORIZED TRANSLATION BY FRED ROTHWELL.	0444264
	B 0079	1509	11 AA SHORT HISTORY OF PHILOSOPHY / CF. J. THONNARD : REV. TRANSLATION FROM THE REV. AND E	0047345
	B 0082	1509	12 AA HISTORY OF PHILOSOPHY: BWITH ESPECIAL REFERENCE TO THE FORMATION AND DEVELOPMENT OF	2901491
	B 0105.65	1509	14 AIHE GREAT CHAIN OF BEING: BA STUDY OF THE HISTORY OF AN IDEA. CIHE WILLIAM JAMES LECIU	4000931
	B 0105.D49	1509	10 ADIALECTICS : BA CONTROVERSY-ORIENTED APPROACH TO THE THEORY OF KNOWLEDGE / CNICHOLAS R	1990057
		1509	10 ADIALECTICS : BA CONTROVERSY-ORIENTED APPROACH TO THE THEORY OF KNOWLEDGE / CNICHOLAS R	3034395
		1509	10 ADIALECTICS : BA CONTROVERSY-ORIENTED APPROACH TO THE THEORY OF KNOWLEDGE / CNICHOLAS R	3034395
		1509	10 ADIALECTICS : BA CONTROVERSY-ORIENTED APPROACH TO THE THEORY OF KNOWLEDGE / CNICHOLAS R	3034395
		1509	10 ADIALECTICS : BA CONTROVERSY-ORIENTED APPROACH TO THE THEORY OF KNOWLEDGE / CNICHOLAS R	3034395
		1509	10 ADIALECTICS : BA CONTROVERSY-ORIENTED APPROACH TO THE THEORY OF KNOWLEDGE / CNICHOLAS R	3034395
	B 0105.D57	1509	13 ALA DISSIM FINIT.	0000341
	B 0105.E7	1509	10 AACTS AND OTHER EVENTS / CJUDITH JARVIS THOMSON.	3074000
		1509	10 AACTS AND OTHER EVENTS / CJUDITH JARVIS THOMSON.	3074000
		1509	10 AACTS AND OTHER EVENTS / CJUDITH JARVIS THOMSON.	3074000
		1509	10 AACTS AND OTHER EVENTS / CJUDITH JARVIS THOMSON.	3074000
	B 0105.E75	1509	11 AEXACT PHILOSOPHY: BPROBLEMS, TOOLS, AND GOALS, CEDITED BY MARIO BUNGE.	0649163
		1509	10 AEXACT PHILOSOPHY: PROBLEMS, TOOLS, AND GOALS, CEDITED BY MARIO BUNGE.	1493055
	B 0105.M8	1509	11 ACIVILT A DELL'UMANESIMO. BATTI DEL VI, VII, VIII CONVEGNO INTERNAZIONALE DEL CENTRO DI	0705030
		1509	11 ADOCTRINES ET FIGURES HUMANISTES.	2000427
		1509	10 AMEDIEVAL FOUNDATIONS OF RENAISSANCE HUMANISM / CWALTER ULLMANN.	3117970
	B 0105.L45	1509	14 AIHE IDEA OF FREEDOM, CBY MORTIMER J. ADLER, FOR THE INSTITUTE FOR PHILOSOPHICAL RESEAR	0727137
		1509	10 AWILL, FREEDOM, AND POWER / CANTHONY KENNY.	4195300
		1509	10 AWILL, FREEDOM, AND POWER / CANTHONY KENNY.	4195300
	B 0105.M4	1509	10 AMEANING, CBY STEPHEN R. SCHIFFER.	0030004
		1509	10 AMEANING AND MORALITY / CCASIMIR LEWY.	4312546

EXAMPLE REPORT 4B (FORMAT 2)

Report 5. List of Book Publishing Record Titles by HEGIS (Instructional Subject)

This report is similar to Report four except that the data are titles, primary subject heading, and classification number of the books published in the U.S. or having a sole U.S. distributer during 1975, 1976, 1977.

The chief purpose of this report is to assist in the evaluation of the LC/HEGIS Translation by revealing titles or subject headings which are inconsistent with the HEGIS categories and to identify titles which have been omitted from categories inadvertently.

The report is interesting in and of itself and might provide a model on which a new service might be provided, i.e. a listing of new publications by HEGIS category for selection purposes or for current awareness services to faculty or other specialists.

The technical problem in segregating the class number from the book number in the BPR file discussed under Report 6/7 is also relevant in Report 9.

EXAMPLE REPORT 5

	0105.A2	150900 NAMES AND DESCRIPTIONS /	REFERENCE (PHILOSOPHY)
	0105.A6	150900 FORMS OF REPRESENTATION /	REPRESENTATION (PHILOSOPHY) CONGRESSES.
	0105.S55	150900 SIMPLICITY /	SIMPLICITY (PHILOSOPHY)
	0105.U5	150900 THE RECONCILIATION OF UNIVERSALS /	UNIVERSALS (PHILOSOPHY)) ADDRESSES, ESSAYS, LECTURES
	0105.U5	150900 UNIVERSALS AND PARTICULARS :	UNIVERSALS (PHILOSOPHY) ADDRESSES, ESSAYS, LECTURES
	0105.W3	150900 IDEOLOGY, REASON, AND THE LIMITATION OF WAR :	JUST WAR DOCTRINE HISTORY.
	0121	150900 A PATH TO ORIENTAL WISDOM :	PHILOSOPHY, ORIENTAL.
	0121.A3	150900 WANDERING IN EDEN :	PHILOSOPHY, ORIENTAL.
	0121.B36	150900 UNDERSTANDING ORIENTAL PHILOSOPHY :	PHILOSOPHY, ORIENTAL HISTORY.
	0121.B36	150900 UNDERSTANDING ORIENTAL PHILOSOPHY :	PHILOSOPHY, ORIENTAL-HISTORY.
	0121.P37	150900 A PATH TO ORIENTAL WISDOM :	PHILOSOPHY, ORIENTAL.
	0126.C43	150900 THE STUDY OF CHINESE PHILOSOPHY /	PHILOSOPHY, CHINESE.
	0126.F5	150900 THE MENCIUS CONCEPTION OF THE CHINESE /	PHILOSOPHY, CHINESE.
	0126.M3	150900 INITIATION TO CHINESE PHILOSOPHY!	PHILOSOPHY, CHINESE.
	0126.SC13	150900 INDIVIDUAL AND STATE IN ANCIENT CHINA :	PHILOSOPHY, CHINESE.
	0127.N4	150900 THE DEVELOPMENT OF NEO-CONFUCIAN THOUGHT /	NEO-CONFUCIANISM.
	0127.T3	150900 THE TAO IS SILENT /	TAO.
	0128.C5	150900 WHY CONFUCIUS HAS BEEN REVERENCED AS THE MODEL TEA	
	0128.M56	150900 MS DN Y UEH (A.D. 168-209) :	
	0128.M58	150900 MSIN-LUN (LUN INSANITY), AND OTHER WRITINGS BY HUA	
	0128.W3	150900 WANG HSIANG-SHAN :	PHILOSOPHY, CHINESE.
	0130	150900 INITIATION TO INDIAN PHILOSOPHY	PHILOSOPHY, INDIC.
	0131.A3B	150900 THE SPIRIT OF INDIAN PHILOSOPHY /	PHILOSOPHY, INDIC.
	0131.C5H7	150900 CONCEPTS OF INDIAN PHILOSOPHY /	PHILOSOPHY, INDIC.
	0131.H66	150900 THE MIND OF INDIA /	PHILOSOPHY, INDIC HISTORY.
	0131.M66	150900 AN INTRODUCTION TO INDIAN THOUGHT /	PHILOSOPHY, INDIC.
	0131.I3	150900 INDIAN THOUGHT :	PHILOSOPHY, INDIC ADDRESSES, ESSAYS, LECTURES.
	0131.I73	150900 WESTERN APPROACHES TO EASTERN PHILOSOPHY /	PHILOSOPHY, INDIC ADDRESSES, ESSAYS, LECTURES.
	0131.M56	150900 FUNDAMENTALS OF INDIAN PHILOSOPHY /	PHILOSOPHY, INDIC HISTORY.
	0132	150900 GEMS FROM THE GITA.	BHAGAVADGITA.
	0132	150900 JOY OF LIFE THROUGH YOGA.	YOGA.
	0132.A8	150900 THE PATH OF SOUL LIBERATION /	ATMAN.
	0132.I3	150900 INDIAN IDEALISM :	IDEALISM.
	0132.L6	150900 MATERIALISM IN INDIAN THOUGHT.	LOK AYATA.
	0132.V3	150900 SEARCH FOR THE ABSOLUTE IN NEO-VEDANTA /	VEDANTA.
	0132.V3	150900 THE FLAME AND THE LIGHT :	VEDANTA COMPARATIVE STUDIES.
	0132.Y6	150900 THE HATHA-YOGA WAY TO HEALTH AND LONGER LIFE.	YOGA.
	0132.Y6	150900 YOGA VEDANTA DICTIONARY	YOGA DICTIONARIES.
	0132.Y6	150900 FOREVER YOUNG, FOREVER HEALTHY.	
	0132.Y6	150900 THE UMAPATI SAMHITA /	YOGA, HA THA.
	0132.Y6	150900 AN INTRODUCTION TO THE YOGA PHILOSOPHY /	YOGA.
	0132.Y6	150900 YOGA AND DESTINY /	YOGA.
	0132.Y6	150900 PATANJALI AND YOGA /	YOGA.
	0132.Y6	150900 GOAL YOGA /	YOGA.
	0132.Y6	150900 YOGA AND THE HINDU TRADITION /	YOGA.
	0132.Y6	150900 COMPLETE HATHA YOGA /	YOGA, HA THA.
	0132.Y6	150900 INITIATION INTO YOGA :	YOGA.
	0132.Y6	150900 YOGA AND PSYCHOTHERAPY :	YOGA.
	0132.Y6	150900 PERSPECTIVES IN YOGA :	YOGA CONGRESSES.
	0132.Y6	150900 SHRI SHRI YOGA :	YOGA.
	0132.Y6	150900 THE DAY WITH YOGA :	YOGA QUOTATIONS, MAXIMS ETC.
	0132.Y6	150900 YOGA DAN SANA :	YOGA.
	0132.Y6	150900 LIVING YOGA :	YOGA ADDRESSES, ESSAYS, LECTURES.
	0136.A8	150900 STUDIES IN RELIGIOUS PHILOSOPHY AND MYSTICISM /	PHILOSOPHY, JEWISH ADDRESSES, ESSAYS, LECTURES.
	0136.J6C	150900 JEWISH PHILOSOPHERS /	PHILOSOPHY, JEWISH HISTORY.
	0162.3000.96P	150900 JAIN PHILOSOPHY :	PHILOSOPHY, JAINA.
	0162.A8	150900 BUDDHIST PHILOSOPHY IN INDIA AND CEYLON /	PHILOSOPHY, BUDDHIST.
	0163.A2	150900 THE MISSION OF GREECE :	PHILOSOPHY, ANCIENT.
	0171.A38	150900 PERMALAION :	PHILOSOPHY, ANCIENT ADDRESSES, ESSAYS, LECTURES.
	0171.A7	150900 PLATO'S UNIVERSE /	COSMOLOGY HISTORY.
	0187.A35	150900 INTELLECTUAL EXPERIMENTS OF THE GREEK ENLIGHTENMEN	PHILOSOPHY, ANCIENT.

The implications of this information for building subject collections are that there is much overlap in the materials and supporting the social sciences and purchasing material in those subjects tends to render multiple benefits. It will be necessary to develop a set of subject groupings with relatively equivalent scope in order to test this phenomenon in other subjects. Before using this information to determine that one should buy more materials in subjects with higher overlap, one should determine whether such materials actually receive more use or experience any other benefits which would justify increased allocation.

FILE SOR00L ELABORATION DATE = 09/20/78

CROSSTABULATION OF

SOR00	SOR00-DEFINED MEGIS GROUPINGS	BY N	# OF FOUR DIGIT MEGIS CATEGORIES CONTAIN

N

SOR00	Count / Row PCT / Col PCT / Tot PCT	1	2	3	4	5	6	7	8	9	10	ROW TOTAL
10		930	2140	2347	2232	1520	1478	1057	624	467	95	13282
BIO & HEALTH PRO		6.8	16.2	17.7	16.8	11.4	11.1	8.0	4.7	3.5	0.7	9.9
		2.4	5.4	6.4	15.3	18.8	17.4	31.5	30.1	29.2	24.8	
		0.7	1.6	1.8	1.7	1.1	1.1	0.8	0.5	0.3	0.1	
12		44	2392	1903	1155	441	672	41	2	16	25	6616
BUSINESS AND MGM		0.7	36.2	28.8	17.5	6.7	10.2	0.6	0.0	0.2	0.4	4.9
		0.1	7.2	7.6	7.9	3.5	5.9	1.2	0.1	1.0	6.5	
		0.0	1.8	1.4	0.9	0.3	0.5	0.0	0.0	0.0	0.0	
14		1808	2340	4364	3241	1632	1063	423	406	281	69	15950
EDUCATION		11.3	14.7	27.3	20.3	10.2	6.7	2.7	2.5	1.8	0.4	11.9
		4.7	8.1	17.4	22.3	20.2	12.5	12.6	19.6	17.5	18.0	
		1.4	1.8	3.3	2.4	1.2	0.8	0.3	0.3	0.2	0.1	
16		1839	1366	758	768	99	445	7	0	0	0	5682
FINE & APPL ARTS		32.4	24.1	13.3	13.5	1.7	7.8	0.1	0.0	0.0	0.0	4.3
		4.8	4.1	3.0	5.3	1.2	5.2	0.2	0.0	0.0	0.0	
		1.4	1.0	0.6	0.6	0.1	0.3	0.0	0.0	0.0	0.0	
19		16282	3958	1146	365	140	448	66	6	0	10	22280
LIT & FOR LANG		73.1	17.8	5.1	1.6	0.6	2.0	0.3	0.0	0.0	0.0	16.7
		42.2	12.3	4.6	2.5	1.7	5.3	2.0	0.3	0.0	2.6	
		12.2	2.7	0.9	0.3	0.1	0.3	0.0	0.0	0.0	0.0	
23		1288	203	139	49	80	315	134	28	16	24	2195
MATH & COMP SCI		58.7	9.2	6.3	2.2	3.6	14.3	6.1	1.3	0.8	1.1	1.6
		3.3	0.9	0.6	0.3	1.0	3.7	5.8	1.4	1.0	6.3	
		1.0	0.2	0.1	0.0	0.1	0.2	0.1	0.0	0.0	0.0	
24		3234	2678	2355	903	783	811	649	332	164	32	12421
PHYS SCI & ENGIN		26.0	21.6	19.0	7.3	6.3	6.5	5.2	2.7	1.3	0.3	9.3
		8.4	9.2	10.2	4.2	9.7	9.5	19.3	16.0	13.2	8.4	
		2.4	2.0	1.9	0.7	0.6	0.6	0.5	0.2	0.1	0.0	
COLUMN TOTAL		38487	29012	25044	14551	8064	8504	3388	2073	1602	363	133661
		28.8	21.7	18.7	10.9	6.0	6.4	2.5	1.6	1.2	0.3	100.0

(CONTINUED)

EXAMPLE REPORT 6

Report 7. Average Prices of New Publications by Instructional Subject.

(a) Format 1

(b) Format 2

This report uses data from the Book Publishing Record data base covering 1975, 1976, 1977 imprints. All records are sorted according to HEGIS category as in Report 6, but in this report the price of each book is recorded and the total cost of all books in each instructional subject is divided by the total number of books in each instructional subject to derive the average cost.

Two separate formats are illustrated.

Format 1 reports the three year total titles and average price, and the details for each of three years by individual four-digit HEGIS code and grouped or summarized by two digit HEGIS codes, e.g. 20 Psychology.

Format 2 reports the same data for each four-digit HEGIS code but does not provide the two-digit summaries.

Recall that books are counted only once in each instructional subject but that they may appear in several instructional subjects. The average is accurate for each instructional subject but may not be applied directly to the total books published in Report 6 to derive the total funds required to buy all the books in each instructional subject. Unless one know the extent of duplication among all of the instructional subjects one cannot establish a total prive for an entire collection without allowing considerable duplication.

AVERAGE PRICES OF BOWKER 1975-77 BPR TITLES BY HEGIS

HEGIS CLASS / SUBCLASS	1975-77 TITLES	1975-77 AVERAGE PRICE	1977 TITLES	1977 AVERAGE PRICE	1976 TITLES	1976 AVERAGE PRICE	1975 TITLES	1975 AVERAGE PRICE
1905 CHEMISTRY,GEN-NO BIOCHEM	398.00	20.51	95.00	21.14	144.00	20.89	159.00	19.77
1906 INORGANIC CHEMISTRY	40.00	27.32	10.00	32.10	18.00	23.88	12.00	28.00
1907 ORGANIC CHEMISTRY	126.00	26.55	27.00	26.33	43.00	26.90	58.00	26.39
1908 PHYSICAL CHEMISTRY	239.00	28.51	61.00	26.61	89.00	29.08	89.00	29.10
1909 ANALYTICAL CHEMISTRY	92.00	29.86	31.00	26.93	27.00	31.00	34.00	31.64
1910 PHARMACEUTICAL CHEMISTRY	18.00	24.33	3.00	24.00	9.00	21.77	6.00	28.33
1911 AGRONOMY	446.00	17.54	124.00	15.84	156.00	17.84	166.00	18.53
1913 ATMOSPHERIC SCI METEOROLG	145.00	17.79	60.00	14.75	65.00	18.72	70.00	15.25
1914 GEOLOGY	519.00	19.38	141.00	20.46	187.00	18.80	191.00	19.15
1915 GEOCHEMISTRY	2.00	5.50	1.00	1.00	1.00	1.00	2.00	5.50
1916 GEOPHYSICS & SEISMOLOGY	21.00	19.76	8.00	17.25	6.00	26.00	7.00	17.28
1917 EARTH SCIENCES,GEN	366.00	18.60	105.00	19.71	127.00	19.30	134.00	16.49
1918 PALEONTOLOGY	129.00	21.78	33.00	22.63	45.00	18.59	51.00	24.03
1919 OCEANOGRAPHY	151.00	21.69	40.00	26.87	43.00	20.53	68.00	19.38
1920 METALLURGY	67.00	27.65	16.00	24.68	15.00	34.63	16.00	24.12
1925 QUANTUM MECHANICS	66.00	19.90	18.00	20.66	21.00	23.61	27.00	16.51
1996 WOOD CHEMISTRY	11.00	21.00	4.00	23.50	2.00	21.00	5.00	19.00
1997 POLYMER CHEMISTRY	48.00	31.80	15.00	30.73	14.00	33.14	17.00	31.64
1998 FOOD SCIENCE	1,433.00	26.42	357.00	25.84	557.00	26.92	519.00	26.28
20 PSYCHOLOGY	3,468.00	12.54	1,057.00	12.36	1,286.00	12.76	1,125.00	12.46
2001 PSYCHOLOGY, GENERAL	208.00	12.35	66.00	12.48	69.00	12.89	73.00	11.72
2002 EXPERIMENTAL PSYCHOLOGY	245.00	12.59	93.00	12.08	84.00	12.69	68.00	13.16
2003 CLINICAL PSYCHOLOGY	1,053.00	13.08	333.00	13.03	403.00	13.05	317.00	13.17
2004 PSYCHOLOGY FOR COUNSELNG	260.00	7.90	91.00	8.91	86.00	7.82	83.00	6.89
2005 SOCIAL PSYCHOLOGY	341.00	9.73	90.00	10.37	140.00	10.13	111.00	9.46
2006 PSYCHOMETRICS	220.00	12.18	76.00	13.28	82.00	11.92	62.00	11.16
2007 STATISTICS IN PSYCHOLOGY	246.00	16.45	85.00	16.67	84.00	17.14	77.00	15.45
2008 INDUSTRIAL PSYCHOLOGY	1.00	11.00	1.00	1.00	1.00	1.00	1.00	11.00
2009 DEVELOPMENTAL PSYCHOLOGY	242.00	11.65	84.00	10.35	77.00	13.05	81.00	11.06
2010 PHYSIOLOGICAL PSYCHOLOGY	484.00	16.42	159.00	13.79	160.00	19.33	165.00	16.13
2011 PSYCHOLOGY OF GENDER	64.00	8.17	16.00	8.12	31.00	8.77	17.00	7.11
2012 PSYCHOLOGY OF PERSONALITY	248.00	10.08	79.00	10.55	107.00	10.04	62.00	9.22
2013 PSYCHOLOGY OF ADJUSTMENT	29.00	10.10	8.00	11.62	8.00	8.00	6.00	11.05
2014 PSYCHOLOGY STATISTICS	26.00	13.34	12.00	10.25	9.00	15.00	5.00	17.79
2015 CHILD & ADOLESCENT PSYC	511.00	9.86	168.00	9.77	165.00	9.73	180.00	10.10
2099 PSYCHOLOGY, OTHER	288.00	11.27	51.00	9.55	111.00	10.82	126.00	12.44
21 PUBLIC AFFAIRS & SERVICE	9,021.00	11.18	2,740.00	11.47	3,347.00	11.13	2,934.00	10.99
2102 PUBLIC ADMINISTRATION	1,637.00	13.95	476.00	15.08	586.00	14.15	575.00	12.80
2103 PARKS & RECREATION MGT	2,831.00	9.19	832.00	9.49	1,092.00	8.88	907.00	9.29
2104 SOCIAL WORK & WELFARE SER	4,598.00	11.86	1,388.00	11.68	1,667.00	11.96	1,543.00	11.66
2105 LAW ENFORCE & CORRECTION	807.00	13.37	255.00	14.12	298.00	13.53	249.00	12.40
22 SOCIAL SCIENCES	29,610.00	13.44	8,543.00	13.40	9,543.00	13.74	8,524.00	13.14
2201 SOCIAL SCIENCES, GENERAL	446.00	14.66	119.00	14.71	183.00	16.43	144.00	13.00
2202 ANTHROPOLOGY	2,699.00	15.23	719.00	14.87	1,056.00	16.17	919.00	14.45
2203 ARCHEOLOGY	849.00	16.90	225.00	15.62	324.00	17.82	300.00	15.44
2204 ECONOMICS	6,665.00	14.12	1,833.00	14.05	2,546.00	14.51	2,286.00	13.73
2205 HISTORY	10,326.00	13.36	2,805.00	13.56	4,006.00	13.33	3,715.00	13.25
2206 GEOGRAPHY	1,373.00	14.44	342.00	14.71	526.00	14.19	455.00	14.49
2207 POLITICAL SCI & GOVERN	5,231.00	12.52	872.00	12.60	1,275.00	12.97	1,084.00	11.94

AVERAGE PRICES OF BOWKER 1975-77 HPR TITLES BY HEGIS

HEGIS CLASS / SUBCLASS	1975-77 TITLES	1975-77 AVERAGE PRICE	1977 TITLES	1977 AVERAGE PRICE	1976 TITLES	1976 AVERAGE PRICE	1975 TITLES	1975 AVERAGE PRICE
0414 FORESTRY	276.00	15.63	73.00	17.53	102.00	16.36	101.00	13.52
0415 NATURAL RESOURCES MGT	89.00	16.93	26.00	21.38	33.00	14.84	30.00	15.36
0417 RANGE MANAGEMENT	111.00	12.09	34.00	14.58	42.00	11.26	35.00	10.68
0201 ENVIRONMENTAL DESIGN,GEN	250.00	13.06	76.00	15.50	97.00	11.46	75.00	12.61
0202 ARCHITECTURE	1,178.00	13.08	318.00	14.66	456.00	11.92	404.00	13.14
0203 INTERIOR DESIGN	102.00	12.63	32.00	12.65	39.00	12.79	31.00	12.41
0204 LANDSCAPE ARCHITECTURE	587.00	13.99	158.00	14.91	226.00	14.27	203.00	12.95
0205 URBAN ARCHITECTURE	74.00	8.12	8.00	4.12	36.00	5.58	30.00	12.23
0206 CITY,COMMUNITY&REGNL PLN	362.00	12.97	84.00	14.44	153.00	11.97	125.00	13.21
0305 AFRICAN STUDIES	572.00	14.85	155.00	15.29	217.00	14.61	200.00	14.80
0306 LATIN AMERICAN STUDIES	735.00	13.54	191.00	14.73	277.00	13.59	267.00	12.64
0401 BIOLOGY GENERAL	572.00	15.85	147.00	15.64	194.00	14.89	231.00	16.80
0402 BOTANY GENERAL	39.00	14.17	14.00	13.85	14.00	12.64	11.00	16.54
0403 BACTERIOLOGY	104.00	16.50	22.00	17.63	46.00	15.76	36.00	16.77
0404 PLANT PATHOLOGY	67.00	21.77	24.00	23.70	21.00	24.71	22.00	16.86
0405 PLANT PHARMACOLOGY	27.00	34.11	5.00	27.33	13.00	33.46	11.00	36.72
0406 PLANT PHYSIOLOGY	76.00	26.91	17.00	29.76	30.00	25.83	29.00	24.00
0407 ZOOLOGY,GENERAL	2,063.00	17.94	550.00	17.24	822.00	18.91	691.00	17.33
0408 PATHOLOGY,HUMAN,ANIMAL	1,007.00	15.28	274.00	15.99	432.00	14.19	301.00	16.21
0409 PHARMACOLOGY,HUMAN,ANIMAL	90.00	19.07	25.00	18.56	39.00	22.46	26.00	4.50
0410 PHYSIOLOGY,HUMAN,ANIMAL	761.00	23.60	211.00	23.14	314.00	25.52	236.00	21.73
0411 MICROBIOLOGY	295.00	22.48	77.00	24.11	112.00	22.55	106.00	21.22
0412 ANATOMY	99.00	17.60	23.00	14.34	38.00	18.52	38.00	18.65
0413 HISTOLOGY	17.00	22.70	5.00	14.80	4.00	20.00	8.00	29.00
0414 BIOCHEMISTRY	247.00	29.91	63.00	30.95	109.00	29.31	75.00	29.92
0415 BIOPHYSICS	44.00	17.50	15.00	15.13	14.00	22.42	15.00	15.26

EXAMPLE REPORT 7B FORMAT 2

Report 8/9. Titles Published vs. Student Credit Hours Generated. Projected

Report six/seven is similar to Report one/two except that bibliographic data from the Book Publishing Record data base is substituted for the bibliographic data representing the cataloging of SUNY units. The purpose of this analysis is similar to that of Report 1 in that it allows the collection manager to relate publishing output to academic program offerings.

The same precautions in interpretation are applicable due to the fact that neither librarians nor institutional planners have reviewed the CASA data or the LC/HEGIS translation to assure that the campus programs are accurately represented. Similarly, no causal relationship should be inferred between the program offerings and the publishing output. There is no direct relationship between the books published and what is taught. Later reports will assess the extent to which the U.S. publishing output represents the pool of materials which libraries purchase.

One technical factor which influences the number of titles reported in each instructional subject is related to the way in which the LC/HEGIS Translation tables were constructed. The class numbers but not the book numbers are included in the Translation. Unlike the OCLC Distribution Tapes, the BPR data base does not separate the two numbers, and therefore some book numbers will be read by the computer programs as part of the class number. As a result, extraneous items may be caught by ranges which include a decimal, e.g. the range for Periodicals of the Russian Revoluation, 1917-1921 is DK 265.A1-DK 265.A19. This range might capture class number DK 265 with book numbers (Cutter numbers) .A1-.A19. as well as those with class numbers in this range. Therefore ranges which specify a decimal may have a slightly inflated count of titles.

In future work an algorithm may be developed to identify and segregate the book number from the class number. This would be useful not only in analyzing records for collection development but for retrospective conversion projects.

Report 10. List of Ranges With Number of BPR Titles Caught by Each (Projected)

This report is similar to Report 3 except that the data are titles reported in BPR for 1975, 1976, and 1977. The immediate use for this report is the evaluation of the LC/HEGIS Translation as well as the secondary purposes mentioned regarding Report 3.

Report 11. Overlap Among Four Digit HEGIS Categories (Instructional Subject in OCLC Records. (Proposed)

Each instructional subject is compared with every other instructional subject in terms of the number of titles they hold in common. The resultant numbers measure the overlap among each pair of instructional subjects. It has been suggested that this overlap implies need for duplicate copies since students from several subjects will require access to these materials. It has also been suggested that there will be efficient use of library materials if overlapping subjects are included in a campus curriculum, i.e. a campus may be well-advised to offer subjects which could share the same materials rather than offer diverse and discrete subjects. Of course, campuses must first meet the needs of potential students and provide programs which will meet the needs of the local community before concerning themselves with the potential multiple use of library materials. This consideration may be relevant in a grouping of multiple institutions in a geographic region which mutually agrees to offer a wider range of programs by coordinating efforts.

Simple two-way analysis does not reveal the total picture of overlap. A fuller analysis would be one in which the extent of overlap between each three, four, etc. instructional groupings are displayed. The type of analysis sought falls under the general rubric of cluster analysis or multidimensional scaling. No programs using these techniques are presently available at SUNY Central, but they should be added to the repertoire of the management information system as it becomes operational in order to provide managers with full information as to the extent of overlap within their collections. This type of analysis would also be very helpful in viewing the publishing output (Report 10) because it would assist in judging the budget sums needed to support programs given the level of overlap.

Report 12. Disciplinary Price Index. (Proposed)

Report 12 uses the same data as Report 11 but displays the data in terms of the deviation of prices from a set base year (1975).

As indicated by Reports 13 and 14, there are considerable discrepancies between the body of books published and those bought by libraries. The titles relevant to college and university acquisition may have an average price which differs considerably from the averages reported in Reports 11 and 12.

Future studies should attempt to acquire actual price data either from the acquisitions records of libraries or the records of jobbers serving a library clientele and comparing these with standard publishing data similar to that in the BPR file.

Report 13. Matching Titles Cataloged by Libraries with Title Published and Reported in BPR. Analyzed by Instructional Subject. (Proposed)

The data used in this report are the totality of the items cataloged by the SUNY units described in Reports 1/2 during the period 1977-78 and the totality of the items published in the U.S. by commercial publishers during the period 1975 through 1977 and contained in the BPR file.

The matching is done by comparing the LC card numbers of the library titles with the LC card number of the BPR titles. This element was selected for matching after a number of alternatives were considered such as the ISBN and LC class number. It was determined that at the present the LC card number was more reliably applied to a larger proportion of bibliographic records in both data bases than any of the alternative elements.

There remain a number of circumstances under which a library might have purchased a title and not have its match recorded in this analysis:

1. The title was purchased and cataloged prior to February 28, 1977.

2. The title has been purchased but has not yet been cataloged.

3. The LC card number is absent or inaccurate on either the BPR or library file.

4. The title is outside the parameters of BPR coverage such as foreign titles, publications of societies and academic institutions, government publications, (analyzed in Report 14).

Each library's titles are matched against the BPR files to determine

1. the number and percent of the total number of BPR titles in each instructional subject. These are also reported in Report 6/7.

2. the number and percent of the total library titles which match the BPR titles by LC card number.

3. the proportion of the total BPR titles which are held by the library.

The titles held by the library but not matched to the Bowker BPR file are analyzed in detail in Report 14. The total number of titles in each Instructional subject are repeated from Report 1/2 to provide a point of comparison between the library's total cataloging and the titles which are found in the BPR file. This information will provide an estimate of the proportion of current cataloging which may be listed in the BPR data base.

Report 14. Analysis of Library Titles Which do not Match BPR Records. (Proposed)

This report is a further analysis of the data extracted in Report 13 regarding the library cataloged titles which do not match the records in BPR, 1975 through 1977.

In this report all subjects are aggregated but the libraries are analyzed separately and by type to determine whether specific collection characteristics for the non BPR titles can be identified.

The data elements such as bibliographic level, record type are taken from the MARC format. Imprint dates, language and country of publication are aggregated for ease of interpretation.

One of the chief purposes of this report is to determine the breakdown of the titles which libraries are cataloging and which do not appear in BPR.

Report 15. Regression Analysis of the Relationship between Selected Academic Factors and Proportion of Current Titles Cataloged in a Group of SUNY Colleges of Arts and Science. (Proposed)

The purpose of the regression analysis is the identification of academic factors and their relationship to the proportion of current U.S. publishing output cataloged by a group of SUNY Colleges.

The data includes the OCLC cataloging of 9 of the 12 SUNY Colleges of Arts and Sciences between 1977 and 1978. The cataloged titles are standardized by subject by dividing by the corresponding number of titles in the BPR data base and reported in Report 6/7. The results of Report 13 would demonstrate that current cataloging in this group of libraries represents approximately N% of the titles listed in BPR, 1975 through 1978. Standardizing the cataloged titles by subject neutralies the effect of differing publishing levels on purchasing patterns by subject.

The purpose of regressing a number of academic variables against the standardized total of books cataloged over a two year period is to determine whether any combination of these variables is related to the subject distribution of cataloged (and inferrentially, acquisitions) over a group of similar libraries. Uncontrolled factors are the total acquisitions budget of the library, the size of library staff which might affect the speed with which items are cataloged, and the unit cost of titles by subject which may partially explain purchasing patterns. Serials are also outside the parameters of this report, although they account for a sizeable portion of library acquisitions.

7. SUMMARY AND CONCLUSIONS

The attempt to develop quantitative components which would be included in a proposed acquisition formula was not successful, although the structure of the formula appears to be validated, given its extension to include use data.

The major accomplishments of the project are:-

1. Development of translation tables which express Hegis taxonomy terms as sets of LC class numbers.

2. Use of these tables to compare library acquisition, institutional administrative data and book publishing data, and to treat them as components of a library management information system.

3. Development of computer programs to build and maintain the tables, isolate and analyze the data, and prepare a series of reports which together comprise the library management information system. The programs are sufficiently flexible to produce an extended range of reports.

4. Establishment of a process by which libraries and institutions can analyze their data.

It is concluded that continued development of the library management data system established in this program, specifically through inclusion of library use data, and validation and maintenance of the translation tables, will establish a library acquisition decision structure which, through its increased sensitivity and responsiveness to local conditions, is of more value, or utility, than a blanket, stereotyped, formula.

8. REFERENCES

1. Broad System of Ordering. 2nd revised Draft. La Haye, Federation de Documentation, June, 1977.

2. Buckland, M. K. Book availability and the library user. New York, Pergamon Press, 1975.

3. Chen, Ching-Chih. Applications of Operations Research Models to Libraries. Cambridge, Mass., The MIT Press, 1976.

4. Chismore, W. Dale and Quentin M. Hill. A Classification of Educational Subject Matter. Washington, D. C. National Center for Educational Statistics. 1978.

5. Clapp, V. W. and R. T. Jordan. Quantitative Criteria for Adequacy of Academic Library Collections. College and Research Libraries. 26, 371-80. September 1965.

6. Colorado Association of College and University Presidents. Special Ad-Hoc Committee. "A Hegis Discipline to Dewey Decimal/Library of Congress Crossover Matrix. Analyzing Library collections by Academic discipline." Unpublished 1977.

7. Edelman, Hendrick. 'Selection Methodology in Academic Libraries'. Unpublished paper delivered at ALA-RTSD Preconference on Collection Development. June, 1977.

8. Evans, Glyn T., Roger Gifford and Donald R. Frans. Collection Development Analysis Using OCLC Archival Tapes. Final Report. Albany, New York. Office of Library Services, SUNY Central Administration, 1977 ED 152-299

9. 'Guidelines for the Formulation of Collection Development Policies'. Library Resources and Technical Services 21, 40 - 47, Winter 1977.

10. Halstead, D. Kent.
Statewide Funding for Higher Education
Washington, D. C., U. S. Department of H.E.W., 1974.

11. Huff, Robert A. and Marjorie O. Chandler
A Taxonomy of Instructional Programs in Higher Education
Washington, D. C., National Center for Educational Statistics.
1970. OE - 50164.

12. Hussain, K. M. The Concept of ICLM in A Resource Requirements Prediction Model (RRPM-1): Guide for the Project Manager. Boulder, Co., National Center for Higher Education Management Systems at WICHE, 1971.

13. International Standard Classification of Education. Paris, Division of Statistics on Education, Office of Statistics, UNESCO, March 1976.

14. Kent, Allen, et al. A cost-benefit model of some critical library operations in terms of use of materials: Final Report.
Pittsburgh, University of Pittsburgh, 1977.

80

15. Knight, John J. "Library Collections and Academic Curricula: Quantitative Relationship" College and Research Libraries - July 1973, Volume 34 (4), p. 295-301.

16. Library adopts UNIMARC
Library of Congress Information Bulletin, 37 (38) p. 581, September 28, 1978

17. Lopez, Manuel D. "Guide for beginning bibliographers". Library Resources and Technical Services, 13, 453-479, Fall, 1969.

18. McGrath, William E. Relationships between Subject Characteristics and use of books in a University Library. Unpublished Dissertations. Syracuse University, 1973.

19. Molinis, R. Marvin. The formula approach to library size: an empirical study of its efficiency in evaluating research libraries. College and Research Libraries, 33 (3), 190-196, May 1972.

20. Moss, Charles E. and Gaither, Gerald W.
Formula budgeting: requiem or renaissance? Journal of Higher Education, Sept/Oct 1976, 47(5) p. 543-563.

21. Pierce, Thomas J. "The Economics of Library Acquisitions: A Book Budget Allocation Model for University Libraries." Unpublished dis. University of Notre Dame, 1976.

22. Precis, a Rotated Subject Index System, by Derek Austin and Peter Butcher. London, Council of the British National Bibliography, Ltd. 1969

23. Raper, James E. Jr.
The Serials Project at the Library of the University of North Carolina at Chapel Hill and a Subject Analysis of its Serial Title Holdings. A research paper for the M.S. in L.S. Degree, May, 1969.

24. Saracevic, T. et al. Causes and dynamics of user frustration in an academic library. College and Research Libraries, 38, 7-18, January 1977.

25. State University of New York. University-wide Faculty Senate. Committee on Library Resources. Factors Underlying Library Collection Development. SUNY Faculty Senate Bulletin, June 1973, 10 (4), 5-6.

26. Voigt, Melvin J.
Acquisition rates in University Libraries. College and Research Libraries, 36, 263-271, July, 1975.

27. Wainwright, E. J. and J. E. Dean. Measures of adequacy for Library Collections in Australian Colleges of Advanced Education. Perth, Western Australia Institute of Technology, 1976. (ED 139-398 and ED 139-399)

28. Washington State University. Interinstitutional Committee of Business Officers. A model budget analysis system for program 07, Libraries. Olympia, Washington, The Committee, March 1973.

9. ACKNOWLEDGMENTS

Although the project staff bear the brunt of the effort in a project such as this, their work would not be realizable were it not for the support, visible and invisible, formal and informal, given by many individuals during the life of the project. It is a pleasure to acknowledge this support in the final report.

First, the Advisory Group, the members of which are Dr. Richard Trueswell, University of Massachusetts; Mrs. Barbara Robinson, Metropolitan Washington Council; Mr. Andrew Uszak, R. R. Bowker Company; Mr. Frank Clasquin, F. W. Faxon and Company; Mr. William Dempsey, Rider College; Dr. James Sherman, Educational Management Services, Inc.; Dr. Alex Cicchinelli, State University of New York Central Administration; Mr. Alan Vincelette, State University of New York Central Administration.

They took their responsibilities very seriously and made a major contribution to the development of the project. Some of the matters discussed by the group were controversial, and frankly not resolved by debate. The discussions were themselves extremely valuable, however, and the issues were aired in a questing, amiable way. It follows that some operational decisions were made that could not please all members of the group. For these, the Project Director takes responsibility.

Special mention must be made of the readiness of Frank Clasquin and Andrew Uszak to make available test data from their respective files of publishing data. That in the event only ABPR data were used is more a reflection of the logistics and schedule of the project than reluctance on the part of theproject or Mr. Clasquin. We still hope that as the project and tape services develop, the research program will be expanded to use Faxon serials data. The project is also pleased to thank Robert Asleson, President of R. R. Bowker, for his support.

In addition to the formal members of the Advisory Group, two observers were invited to participate. They were Hendrik Edelman of Cornell, who is administering a Mellon Grant to review collection development policies at Cornell University, and representatives from the Association for Research Libraries Research Library Management Collection Analysis Project (CAP). In the event, the schedules of very busy people did not allow ARL representatives to attend. They have received, however, all project materials. Mr. Edelman brought his considerable experience and insight to the project. He delivered unstintingly from this reservoir in inimitable style.

The final meeting of the advisory group was held at the Library of Congress, where the project is of interest to the Subject Cataloging Division and the Network Development Office. Mary K. Pietris, Head of Subject Cataloging, made interesting and thoughtful contributions to the final meeting.

Dr. Curtis O. Baker, Director of Surveys National Center for Education Statistics, were very hospitable and helpful to project staff, and gave valuable guidance at key stages of the project.

It is important to establish regular liaison with both LC and NCES regarding the maintenance of the tables and the project is encouraged to believe that this will be possible.

The 'invisible college' was, both before and during the project, very tangible to the Director. Many co-workers with whom the project was discussed gave sound advice. Remembering with special pleasure conversations with Barbara Evans Markuson

(Incolsa), Robert Burns (University of Colorado), Miriam Drake (Perdue University), William McGrath (University of Lowell), Alice Wilcox (Minitex) and Steve Salmon (University of California) in no way diminishes the value of all the discussions. One tangible expression of interest was the number of unpublished papers received by the project, a number of which are cited. All the donors are thanked for their support.

The work being performed by Mrs. Virginia Bowden (University Texas Health Science Library, San Antonio) has been cited in the text. It is pleasing to record here the value of our discussions.

Don Franz, consultant to the project, gave sound and constructive advice which clarified some real difficulties in the project's early stage.

The staff support to this project has been unusually diffuse. First, the subject specialists who defined the tables of LC/HEGIS conversion (see Appendix 2). They undertook a difficult and unprecedented task with skill and enthusiasm. The results are not perfect, of course; they never can be. They are, however, sound, worthy of expansions, revision, and maintenance, and for that the project is very grateful.

The extremely arduous and persnickety task of proofreading the tables was undertaken by Debbie Lewis and Janet Dahlquist with skill and patience. A number of unpaid editors and proofreaders were pressed into service to read various tables and texts, including Al Beilby, Don Watson and Teresa Strozik. We thank them.

Making the project work, and processing the myriad bits of paper, is vital. Patricia Heeran, Secretary, and Charles Yandell, Network Business Manager, of the Office of Library Services were knowledgeable and patient, as was Patricia Winters, of the SUNY Research Foundation. The final report was typed and retyped by Esther Lawyer and Maureen Scheely, much of the time from inscrutable handwriting. We are grateful for their skill and patience.

Finally, as Project Director, it is my privilege to acknowledge my co-workers Mary Beilby and Roger Gifford. Their efforts were skillful, prodigious, and productive. The project owes its present state largely to their dedication and enthusiasm, and as Director I'm pleased to record my thanks.

Appendix 1

CASA MODIFIED HEGIS TAXONOMY

THIS IS A LIST OF THE FEDERAL HEGIS TAXONOMY OF INSTRUCTIONAL PROGRAMS IN HIGHER EDUCATION AS MODIFIED TO ACCOMMODATE THE CODING OF INSTRUCTIONAL ACTIVITIES. THE TAXONOMY HAS BEEN EXPANDED TO ALLOW THE SPECIFIC CODING OF COURSES WHICH WOULD OTHERWISE HAVE TO BE SHOWN AS XX99 OTHER, SPECIFY. THIS MODIFIED VERSION OF THE HEGIS TAXONOMY IS NOT USED FOR THE PURPOSE OF REPORTING ENROLLMENT AND DEGREES GRANTED TO THE FEDERAL GOVERNMENT WHICH MUST FOLLOW THE ORIGINAL SPECIFICATION OF THE TAXONOMY.

01 AGRICULTURE&NATURAL RES

0100 AGRICULTURE UNCLASSIFIED
0101 AGRICULTURE,GEN
0102 AGRONOMY,CROP MANAGEMENT
0103 SOILS SCIENCE-MGT&CONSRV
0104 ANIMAL SCIENCE-HUSBANDRY
0105 DAIRY SCIENCE-HUSBANDRY
0106 POULTRY SCIENCE
0107 FISH,GAME,WILDLIFE MGT
0108 HORTICULTURE-FRUIT&VEG
0109 ORN HORT,NURSERY SCIENCE
0110 AGRICULTURE & FARM MGT
0111 AGRICULTURE ECONOMICS
0112 AGRICULTURE BUSINESS
0113 FOOD SCIENCE & TECH
0114 FORESTRY
0115 NATURAL RESOURCES MGT
0116 AGRI&FORESTRY TECH
0117 RANGE MANAGEMENT
0199 AGRICULTURE, OTHER

02 ARCHITECTURE&ENVIRON DES

0200 ARCHITECTURE UNCLASSIFID
0201 ENVIRONMENTAL DESIGN,GEN
0202 ARCHITECTURE
0203 INTERIOR DESIGN
0204 LANDSCAPE ARCHITECTURE
0205 URBAN ARCHITECTURE
0206 CITY,COMMUNITY&REGNL PLN
0299 ARCHITECTURE, OTHER

03 AREA STUDIES

0300 AREA STUDIES UNCLASSIFID
0301 ASIAN STUDIES,GENERAL
0302 EAST ASIAN STUDIES
0303 SOUTH ASIAN STUDIES
0304 SOUTHEAST ASIAN STUDIES
0305 AFRICAN STUDIES
0306 ISLAMIC STUDIES
0307 RUSSIAN & SLAVIC STUDIES
0308 LATIN AMERICAN STUDIES
0309 MIDDLE EASTERN STUDIES
0310 EUROPEAN STUDIES, GENERL
0311 EASTERN EUROPEAN STUDIES
0312 WEST EUROPEAN STUDIES
0313 AMERICAN STUDIES
0314 PACIFIC AREA STUDIES
0395 CANADIAN STUDIES
0396 PUERTO RICAN STUDIES
0397 JUDAIC STUDIES
0398 SW ASIAN & N AFRICAN STD
0399 AREA STUDIES, OTHER

04 BIOLOGICAL SCIENCES

0400 BIOLOGICAL SCI UNCLASSFD
0401 BIOLOGY GENERAL
0402 BOTANY GENERAL
0403 BACTERIOLOGY
0404 PLANT PATHOLOGY
0405 PLANT PHARMACOLOGY
0406 PLANT PHYSIOLOGY
0407 ZOOLOGY,GENERAL
0408 PATHOLOGY,HUMAN,ANIMAL
0409 PHARMCOLOGY,HUMAN,ANIMAL
0410 PHYSIOLOGY,HUMAN,ANIMAL
0411 MICROBIOLOGY
0412 ANATOMY
0413 HISTOLOGY
0414 BIOCHEMISTRY
0415 BIOPHYSICS
0416 MOLECULAR BIOLOGY
0417 CELL BIOLOGY,CYTOLOGY

0418 MARINE BIOLOGY
0419 BIOMETRICS &BIOSTATISTCS
0420 ECOLOGY
0421 ENTOMOLOGY
0422 GENETICS
0423 RADIOBIOLOGY
0424 NUTRITION,SCIENTIFIC
0425 NEUROSCIENCES
0426 TOXICOLOGY
0427 EMBRYOLOGY
0489 REGULATORY BIOLOGY
0490 BIOLOGY OF SEX
0491 PARASITOLOGY
0492 ETHNOLOGY
0493 MYCOLOGY
0494 TAXONOMY
0495 FOREST ZOOLOGY
0496 FOREST ENTOMOLOGY
0497 FOREST BOTANY
0498 FOREST BIOLOGY
0499 BIOLOGY,OTHER

05 BUSINESS & MANAGEMENT

0500 BUSINESS,UNCLASSIFIED
0501 BUSINESS & COMMERCE, GEN
0502 ACCOUNTING
0503 BUSINESS STATISTICS
0504 BANKING & FINANCE
0505 INVESTMENTS & SECURITIES
0506 BUSINESS MGT & ADMINISTR
0507 OPERATIONS RESEARCH
0508 HOTEL & RESTAURANT MGT
0509 MARKETING & PURCHASING
0510 TRANSPORTATION&PUBLIC UT
0511 REAL ESTATE
0512 INSURANCE
0513 INTERNATIONAL BUSINESS
0514 SECRETARIAL STUDY-BA&HGR
0515 PERSONNEL MANAGEMENT
0516 LABOR & INDUSTRIAL REL
0517 BUSINESS ECONOMICS
0596 BUSINESS LAW
0597 ORGANIZATIONAL BEHAVIOR
0598 REGIONAL SCIENCE
0599 OTHER, SPECIFY

06 COMMUNICATIONS

0600 COMMUNICATIONS UNCLASSFD
0601 COMMUNICATIONS, GENERAL
0602 JOURNALISM-PRINTED MEDIA
0603 RADIO/TV
0604 ADVERTISING
0605 COMMUNICATION,MEDIA
0699 COMMUNICATIONS, OTHER

07 COMPUTER & INFO SCIENCE

0700 COMPUTER SCI UNCLASSIFD
0701 COMPUTER & INFO SCI,GEN
0702 INFO SCIENCE & SYSTEMS
0703 DATA PROCESSING
0704 COMPUTER PROGRAMMING
0705 SYSTEMS ANALYSIS
0798 COMPUTER STRUCT & DESIGN
0799 COMPUTER SCI, OTHER

08 EDUCATION

0800 EDUCATION UNCLASSIFIED
0801 EDUCATION,GENERAL
0802 ELEMENTARY EDUCATION,GEN
0803 SECONDARY EDUCATION,GEN
0804 JUNIOR HIGH SCHOOL ED
0805 HIGHER EDUCATION,GEN
0806 JR & COMMUNITY COLL ED
0807 ADULT & CONTINUING ED
0808 SPECIAL EDUCATION
0809 ADMINISTRATION SPEC ED
0810 ED OF MENTALLY RETARDED
0811 EDUCATION OF THE GIFTED
0812 EDUCATION OF THE DEAF
0813 ED OF CULTURALLY DISADV
0814 ED OF VISUALLY HANDICAPD
0815 SPEECH CORRECTION
0816 ED OF EMOTIONALLY DISTRB
0817 REMEDIAL EDUCATION
0818 SPEC LEARNING DISABLTS
0819 ED OF PHYSICALLY HANDCPD
0820 ED OF MULTIPLE HANDICAPD
0821 SOCIAL FOUNDATIONS
0822 EDUCATIONAL PSYCHOLOGY
0823 PRE-ELEMENTARY ED-KNDGDN
0824 ED STATISTICS & RESEARCH
0825 ED TESTING,EVAL,MEASURES
0826 STUD PERSONNEL-COUNSELNG
0827 EDUCATIONAL ADMIN
0828 EDUCATIONAL SUPERVISION
0829 CURRICULUM & INSTRUCTION
0830 READING EDUCATION
0831 ART EDUCATION
0832 MUSIC EDUCATION
0833 MATHEMATICS EDUCATION
0834 SCIENCE EDUCATION
0835 PHYSICAL ED METHODOLOGY
0836 DRIVER & SAFETY ED
0837 HEALTH & FAMILY LIFE ED

0838 BUSINESS & COMMERCE ED
0839 INDUSTRIAL ART,VOC TECH
0891 BILINGUAL EDUCATION
0892 HOME ECONOMICS EDUCATION
0893 ENGLISH EDUCATION
0894 SOCIAL STUDIES EDUCATION
0895 EDUCATIONAL COMMUNICATN
0896 SPEECH EDUCATION
0897 FOREIGN LANG,SECONDRY SH
0898 PHYS ED-GYM CLASSES
0899 EDUCATION, OTHER

09 ENGINEERING

0900 ENGINEERING UNCLASSIFIED
0901 ENGINEERING, GENERAL
0902 AEROSPACE, ENGINEERING
0903 AGRICULTURAL ENGINEERING
0904 ARCHITECTURAL ENGINEERING
0905 BIO- BIOMEDICAL ENGINRNG
0906 CHEMICAL ENG-PETRCL REFN
0907 PETROLEUM ENG-NOT REFINE
0908 CIVIL,CONSTR,TRANSP,ENG
0909 ELECTRICAL,COMMUNICA ENG
0910 MECHANICAL ENGINEERING
0911 GEOLOGICAL ENGINEERING
0912 GEOPHYSICAL ENGINEERING
0913 INDUSTRIAL MANAGEMT ENG
0914 METALLURGICAL ENGINEERG
0915 MATERIALS ENGINEERING
0916 CERAMIC ENGINEERING
0917 TEXTILE ENGINEERING
0918 MINING&MINERAL ENG
0919 ENGINEERING PHYSICS
0920 NUCLEAR ENGINEERING
0921 ENGINEERING MECHANICS
0922 ENVIRONMENT&SANITARY ENG
0923 NAVAL ARCHTEC&MARINE ENG
0924 OCEAN ENGINEERING
0925 ENGINEERING TECHNOLOGIES
0997 PAPER SCI,&COD PROD ENG
0998 FOREST ENGINEERING
0999 ENGINEERING, OTHER

10 FINE AND APPLIED ARTS

1000 FINE ARTS UNCLASSIFIED
1001 FINE ARTS, GENERAL
1002 ART-PAINTING,DRAW&SCULPT
1003 ART HISTORY&APPRECIATION
1004 MUSIC PERFORM,COMP,THEOR
1005 MUSIC-LIBERAL ARTS PROG
1006 MUSIC HISTORY&APPRECIATN
1007 DRAMATIC ARTS
1008 DANCE
1009 APPLIED DESIGN
1010 CINEMATOGRAPHY
1011 PHOTOGRAPHY
1094 PUPPETRY
1095 MIME AND MOVEMENT
1096 FILM HISTORY AND APPREC
1097 CIRCUS ARTS
1098 ART CONSERVATION
1099 FINE ARTS,OTHER

11 FOREIGN LANGUAGES

1100 FOREIGN LANG UNCLASSIFD
1101 FOREIGN LANGUAGES, GEN
1102 FRENCH
1103 GERMAN
1104 ITALIAN
1105 SPANISH
1106 RUSSIAN
1107 CHINESE
1108 JAPANESE
1109 LATIN
1110 GREEK,CLASSICAL
1111 HEBREW & SEMITIC
1112 ARABIC
1113 INDIAN-ASIATIC
1114 SCANDINAVIAN LANGUAGES
1115 SLAVIC LANG,EXCEPT RUSSN
1116 AFRICAN LANGUAGES
1190 INDONESIAN
1191 HUNGARIAN
1192 AMHARIC
1193 PORTUGUESE
1194 TURKISH
1195 GREEK,MODERN
1196 PERSIAN
1197 SEMITIC LANGUAGES
1198 SANSKRIT
1199 FOREIGN LANGUAGES, OTHER

12 HEALTH PROFESSIONS

1200 HEALTH PROF UNCLASSIFD
1201 HEALTH PROFESSIONS, GEN
1202 HOSPITAL&HEALTH CARE ADM
1203 NURSING-BA & HIGHER
1204 DENTISTRY,DDS,DMD DEGREE
1205 DENTAL SPECIALITIES
1206 MEDICINE,MD DEGREE
1207 MEDICAL SPECIALITES
1208 OCCUPATIONAL THERAPY
1209 OPTOMETRY
1210 OSTEOPATHIC MEDICINE

1211 PHARMACY
1212 PHYSICAL THERAPY
1213 DENTAL HYGIENE-BA&HIGHER
1214 PUBLIC HEALTH
1215 MEDICAL RECORD LIBRARIAN
1216 PODIATRY, PODIATRIC MED
1217 BIOMEDICAL COMMUNICATNS
1218 VETERINARY MEDICINE
1219 VETERINARY MEDICINE SPEC
1220 SPEECH PATHOLGY&AUDIOLGY
1221 CHIROPRACTIC
1222 CLINICAL SOCIAL WORK
1223 MEDICAL LAB TECH-BA&HIGH
1224 DENTAL TECH - BA&HIGHER
1225 RADIOLOGIC TECH-BA&HIGH
1229 HEALTH PROF, OTHER
1298 FIRST AID & SAFETY
1299 HEALTH PROFESSIONS, OTHER

13 HOME ECONOMICS

1300 HOME EC UNCLASSIFIED
1301 HOME ECONOMICS, GENERAL
1302 HOME DECORATION & EQUIP
1303 CLOTHING & TEXTILES
1304 CONSUMER ECO, HOME MGTMT
1305 FAMILY RELATIONS
1306 FOODS&NUTRITION-DIETICS
1307 INSTITUTNL-CAFETRIA MGT
1399 HOME EC, OTHER

14 LAW

1400 LAW UNCLASSIFIED
1401 LAW, GENERAL
1499 LAW, OTHER

15 LETTERS

1500 LETTERS UNCLASSIFIED
1501 ENGLISH, GENERAL
1502 LITERATURE, ENGLISH
1503 COMPARATIVE LITERATURE
1504 CLASSICS
1505 LINGUISTICS
1506 SPEECH, DEBATE, FORENSIC
1507 CREATIVE WRITING
1508 TEACHING ENGLISH AS FORL
1509 PHILOSOPHY
1510 RELIGIOUS STUDIES
1593 TECHNICAL WRITING
1594 DEVELOPMENTAL READING
1595 CHILDRENS' LITERATURE
1596 LITERATURE OF THE STAGE
1597 AMERICAN LITERATURE
1598 MYTHOLOGY
1599 LETTERS, OTHER

16 LIBRARY SCIENCE

1600 LIBRARY SCI UNCLASSIFIED
1601 LIBRARY SCIENCE, GENERAL
1699 LIBRARY SCI, OTHER

17 MATHEMATICS

1700 MATHEMATICS UNCLASSIFIED
1701 MATHEMATICS, GENERAL
1702 STATISTICS
1703 APPLIED MATHEMATICS
1799 MATHEMATICS, OTHER

18 MILITARY SCIENCE

1800 MILITARY SCI UNCLASSIFD
1801 MILITARY SCIENCE-ARMY
1802 NAVAL SCI-NAVY&MARINES
1803 AEROSPACE SCI-AIR FORCE
1899 MILITARY SCI, OTHER

19 PHYSICAL SCIENCES

1900 PHYSICAL SCI UNCLASSIFD
1901 PHYSICAL SCIENCES, GEN
1902 PHYSICS, GEN-NO BIOPHYSCS
1903 MOLECULAR PHYSICS
1904 NUCLEAR PHYSICS
1905 CHEMISTRY, GEN-NO BIOCHEM
1906 INORGANIC CHEMISTRY
1907 ORGANIC CHEMISTRY
1908 PHYSICAL CHEMISTRY
1909 ANALYTICAL CHEMISTRY
1910 PHARMACEUTICAL CHEMISTRY
1911 ASTRONOMY
1912 ASTROPHYSICS
1913 ATMOSPHERIC SCI METEORLG
1914 GEOLOGY
1915 GEOCHEMISTRY
1916 GEOPHYSICS & SEISMOLOGY
1917 EARTH SCIENCES, GEN
1918 PALEONTOLOGY
1919 OCEANOGRAPHY
1920 METALLURGY
1995 QUANTUM MECHANICS
1996 WOOD CHEMISTRY
1997 POLYMER CHEMISTRY
1998 FOREST CHEMISTRY
1999 PHYSICAL SCI, OTHER

87

20 PSYCHOLOGY

2000 PSYCHOLOGY UNCLASSIFIED
2001 PSYCHOLOGY, GENERAL
2002 EXPERIMENTAL PSYCHOLOGY
2003 CLINICAL PSYCHOLOGY
2004 PSYCHOLOGY FOR COUNSELNG
2005 SOCIAL PSYCHOLOGY
2006 PSYCHOMETRICS
2007 STATISTICS IN PSYCHOLOGY
2008 INDUSTRIAL PSYCHOLOGY
2009 DEVELOPMENTAL PSYCHOLOGY
2010 PHYSIOLOGICAL PSYCHOLOGY
2093 PSYCHOLOGY OF GENDER
2094 PSYCHOLOGY OF QUEST SYMB
2095 PSYCHOLOGY OF PERSONALTY
2096 PSYCHOLOGY OF ADJUSTMENT
2097 PSYCHO LINGUISTICS
2098 CHILD & ADOLESCENT PSYC
2099 PSYCHOLOGY, OTHER

21 PUBLIC AFFAIRS & SERVICE

2100 PUBLIC AFFAIRS UNCLASSFD
2101 COMMUNITY SERVICES, GEN
2102 PUBLIC ADMINISTRATION
2103 PARKS & RECREATION MGT
2104 SOCIAL WORK & HELPNG SER
2105 LAW ENFORC & CORRECTION
2106 INTERNATIONAL PUBLIC SER
2199 PUBLIC AFFAIRS, OTHER

22 SOCIAL SCIENCES

2200 SOCIAL SCI UNCLASSIFIED
2201 SOCIAL SCIENCES, GENERAL
2202 ANTHROPOLOGY
2203 ARCHEOLOGY
2204 ECONOMICS
2205 HISTORY
2206 GEOGRAPHY
2207 POLITICAL SCI & GOVERN
2208 SOCIOLOGY
2209 CRIMINOLOGY
2210 INTERNATIONAL RELATIONS
2211 AFRO-AMERICAN STUDIES
2212 AMERICAN INDIAN CULTURE
2213 MEXICAN-AMERICAN CULTURE
2214 URBAN STUDIES
2215 DEMOGRAPHY
2290 BEHAVIORAL SCIENCE
2297 MINORITY STUDIES
2298 WOMENS STUDIES
2299 SOCIAL SCIENCES, OTHER

24 FORESTRY

2400 FORESTRY UNCLASSIFIED
2401 FORESTRY,GENERAL
2402 FOREST SOILS
2404 FOREST ECONOMICS
2405 RESOURCES MANAGEMENT
2406 FOREST MANAGEMENT
2407 FOREST RECREATION
2408 SILVICULTURE
2409 WORLD FORESTRY
2411 SOIL & WATER CONSERV
2412 FOREST RESEARCH
2499 FORESTRY, OTHER

49 INTERDISCIPLINARY STUDY

4900 INTERDISC UNCLASSIFIED
4901 GENERAL LIBERAL ARTS&SCI
4902 BIOLOGICAL & PHYSICAL SC
4903 HUMANITIES & SOCIAL SCI
4904 ENGINEERING & OTHER DISC
4994 ENVIRONMENTAL STUDIES
4996 WATER RESOURCES
4997 ORGANIC MATERIALS SCI
4998 ORIENTATION
4999 INTERDISCIPLINARY, OTHER

50 BUSINESS & COMMERCE TECH

5000 BUSINESS TECH UNCLASSIFD
5001 BUS & COMMERCE TECH,GEN
5002 ACCOUNTING TECHNOLOGIES
5003 BANKING & FINANCE TECH
5004 MARKETING,DISTRIBUTION
5005 SECRETARIAL TECHNOLOGIES
5006 PERSONAL SERVICE TECH
5007 PHOTOGRAPHY TECHNOLOGIES
5008 COMMUNICATNS&BRDCST TECH
5009 PRINTING & LITHOGRAPHY
5010 HOTEL & RESTAURANT MGT
5011 TRANSPORT&PUBLIC UTILS
5012 APPLIED ARTS,GRAPHIC ART
5098 VOCATIONAL CLERICAL SKLS
5099 BUSINESS TECH, OTHER

51 DATA PROCESSING TECH

5100 DATA PROCESS TECH UNCLAS
5101 DATA PROCESSING TECH,GEN
5102 KEY PUNCH OPERATOR TECH
5103 COMPUTER PROGRAMMER TECH
5104 COMP OPR & PERIPHERAL
5105 DATA PROCESSING EQUIP MA
5199 DATA PROCESS TECH, OTHER

52 HEALTH SERVICE&PARAMED

5200 HEALTH SERVICE UNCLASSFD
5201 HEALTH SERVICES ASST TEC
5202 DENTAL ASST TECH
5203 DENTAL HYGIENE TECH
5204 DENTAL LAB TECH
5205 MED OR BIO LAB ASST TECH
5206 ANIMAL LAB ASST TECH
5207 RADIOLOGIC TECHNOLOGIES
5208 NURSING,RN-LESS THAN 4YR
5209 NURSING,PRACTICAL-LT-4YR
5210 OCCUPATIONAL THERAPY TEC
5211 SURGICAL TECHNOLOGIES
5212 OPTCAL TECHNOLOGIES
5213 MEDICAL RECORD TECHNOLOG
5214 MEDICAL ASST & MED OFC-A
5215 INHALATION THERAPY TECH
5216 PSYCHIATRIC TECH
5217 ELECTRC DIAGNOSTIC TECH
5218 INSTITUTIONAL MGT TECH
5219 PHYSICAL THERAPY TECH
5297 ORTHOTICS,ARTIFICIAL LMB
5298 MORTUARY SCIENCE
5299 HEALTH SERVICE, OTHER

53 MECHANICAL & ENGINR TECH

5300 MECHANICAL TECH UNCLASFD
5301 MECH & ENG TECH, GENERAL
5302 AERONAUTICAL & AVIA TECH
5303 ENGINEERING GRAPHICS
5304 ARCHITECTURAL DRAFTNG TC
5305 CHEMICAL TECHNOLOGIES
5306 AUTOMOTIVE TECHNOLOGIES
5307 DIESEL TECHNOLOGIES
5308 WELDING TECHNOLOGIES
5309 CIVIL TECHNOLOGIES
5310 ELECTRONICS & MACHINE TC
5311 ELECTROMECHANICAL TECH
5312 INDUSTRIAL TECHNOLOGIES
5313 TEXTILE TECHNOLOGIES
5314 INSTRUMENTATION TECH
5315 MECHANICAL TECHNOLOGIES
5316 NUCLEAR TECHNOLOGIES
5317 CONSTRUCTION & BUILDING
5393 AGRICULTURAL ENGINEERING
5394 AIR CONDITIONING TECH
5395 ELECTRICAL TECHNOLOGY
5396 WOOD UTILIZATION TECH
5397 VOCATIONAL GARAGE MECH
5398 PLASTIC TECHNOLOGY
5399 MECHANICAL TECH, OTHER

54 NATURAL SCIENCE TECH

5400 NATURAL SCI TECH UNCLSFD
5401 NATURAL SCIENCE TECH,GEN
5402 AGRICULTURE TECHNOLOGIES
5403 FORESTRY & WILDLIFE TECH
5404 FOOD SERVICES TECHNOLOGS
5405 HOME ECONOMICS TECHNLGS
5406 MARINE & OCEANOGRAPHIC T
5407 LABORATORY TECHNOLOGS,GN
5408 SANITATION & PUBLIC HLTH
5493 FOOD PROCESSING TECHNOLG
5494 DAIRY & FOOD SCIENCE
5495 AGRICULTURAL ENGINEERING
5496 AGRONOMY
5497 AGRICULTURAL BUSINESS
5498 ANIMAL HUSBANDRY
5499 NATURAL SCI TECH, OTHER

55 PUBLIC SERVICE RELATED

5500 PUBLIC SERV TECH,UNCLSFD
5501 PUBLIC SERV TECH, GEN
5502 BIBLE STUDY OR RELIGION
5503 EDUCATION TECHNOLOGIES
5504 LIBRARY ASSISTANT TECH
5505 LAW ENFORCE, CORRECTIONS
5506 RECREATION&SOCIAL WK TCH
5507 FIRE CONTROL TECHNOLOGY
5508 PUBLIC ADMINISTRATION TC
5597 CHILD CARE
5598 PLANNING TECHNOLOGY
5599 PUBLIC SERVICE, OTHER

Appendix 2

Agriculture and Natural Resources	David Thomas Mann Library College of Environmental and Life Sciences Cornell University
Architecture and Environmental Design	Mary Beilby Research Team Coordinator
Area Studies African Studies	Mark R. Yerburgh University Library SUNY at Albany
Russian and Slavic Studies	Kay Schaffer University Library SUNY at Albany
Latin American Studies	Dan C. Hazen Cornell University Libraries
Biological Sciences	Mary Beilby Research Team Coordinator
Forestry	Katherine Rossi, Salvacion Dela Paz, Dianne Capritta SUNY College of Environmental Science and Forestry Syracuse, New York
Business and Management	Ms. Sue Bollman Management Librarian Rush Rhees Library University of Rochester
Communications	Mary Beilby
Computer and Information Science	Steve Rockey Mathematics Librarian Cornell University
Education	Mary Beilby
Special Education	Mary Hall Butler Library State University College at Buffalo

Psychology	Mary Beilby
Public Affairs and Service	
Public Administration	Mark Yerburgh University Library SUNY at Albany
Parks and Recreation Management	Mary Beilby
Social Work and Helping Services	David Carr Alexander Library Rutgers University
Law Enforcement and Correction	Theresa Capone John Jay College of Criminal Justice
Social Sciences	
Anthropology Archeology	Mary Beilby
History	William Weiss Russell Sage College Library Troy, New York
Geography	Mary Beilby
Political Science and Government	Thomas L. Bonn Political Science Bibliography SUNY at Cortland
Sociology	Mary Beilby
Criminology	Theresa Capone John Jay College of Criminal Justice
International Relations	Thomas L. Bonn Political Science Bibliography SUNY at Cortland
Afro-American Studies	Mark R. Yerburgh University Library SUNY at Albany
Urban Studies Demography	Mary Beilby
Women's Studies	Mina B. LaCroix University Library SUNY at Albany

Appendix 3

INSTRUCTIONS FOR THE ASSIGNMENT OF L.C. TO HEGIS CATEGORIES

PURPOSE OF THE PROJECT

The purpose of relating the two classification schemes is to identify which L.C. classified library materials are supporting which university programs (classified by HEGIS). The resulting list of HEGIS categories with L.C. numbers will be used to analyze library collections and current publishing output in relation to the academic programs.

Although the primary analyses anticipated are those which include all of the library materials related to an academic program there may be analyses requiring subsets of a program. In order to analyze subsets these need to be specified initially.

PROCEDURE

You are to begin with the four-digit HEGIS categories and treat them as if they were the titles of academic programs. You should then bring to bear your knowledge of the subject--listing the subjects which are typically included in an academic program. This listing should not be specific to any one institution but be as general as possible. Some ideas of what to include can be gleaned from talks with other experts in the discipline, abstract/index journals, bibliographies, college catalogs, L.C. Class schedules and other sources you choose.

Overlap. The overlapping nature of disciplines indicates that the same library materials will be used for several academic programs. These overlaps should be included in your work. However, we want to avoid drawing the entire classification into each academic program. For instance, administration is taught in an education program. Include in education those numbers which refer specifically to educational administration but not all of public administration or business administration. Include educational psychology but not all of personality theory. In a medical program include biomedical communication classifications but not all of communication theory.

Level of detail. It is conceivable that each HEGIS category could be given a single L.C. number or range of numbers. For example: sociology could simply be described as HM-HX. However, this level of analysis is not specific enough to enable one to analyze major subjects within sociology such as sociolinguistics. Therefore, it is desirable to utilize some of the major groupings within the L.C. classification to provide a logical set of subjects within sociology. The problem is how far to go in breaking down major subjects, and no definite solution can be prescribed. The general guideline of between 10 and 50 is offered, but this will be modified by the judgment of the subject specialist.

NOTES ON THE APPLICATION OF HEGIS

HEGIS "GENERAL" is for subjects too broad for the specific HEGIS categories. The tendancy might be to consider all of a discipline's subject matter appropriate to a

general course in a subject. We want to limit the use of "general", however, to broad topics such as natural history, general biology, etc. which are found at the beginning of the classification.

HEGIS "UNCLASSIFIED" is used for thesis research and other undefined courses; therefore "UNCLASSIFIED" should be ignored in this project.

HEGIS "OTHER" is used for specific topics not defined by the HEGIS taxonomy. It is similar to "general specific" in L.C.

CODING HEGIS/LC RANGES

There are three parts in each HEGIS/LC range:

1. Four digit HEGIS number (column 7 to 10)
2. LC Classification number (column 11 to 35)
3. Text description of the range (column 36 to 80)

Please block letter on the coding forms provided or type in three distinct columns on plain typing paper. Type along the eleven inch side of the paper.

Distinguish the following ambiguous symbols in the following way:

O = alpha O	Ø = zero
2 = two	Z = alpha z
1 = one	I = alpha i

Details regarding the coding of ranges:

1. Enter the four digit HEGIS number which you are working with. If the same HEGIS number is repeated draw an arrow and brackets to avoid rewriting the number.

   2208

   2207
   2204
   2207

2. L.C. Classification number. Code the L.C. Classification number as either a single number or as two numbers separated by a hyphen.

   (a) HA 3
   (b) HA 3-HA 3
   (c) HA 3-HA 5

   The single number in example (a) will allow only records with that precise number to be analyzed. The range in example (b) will allow all decimal subdivisions of HA 3 such as HA 3.6 or HA 3.C5 to be included. The third example will allow all of HA 3, HA 4, and HA 5 including their decimal subdivisions to be included.

The computer programs work only with the classification numbers derived from the classification schedules and not with book numbers assigned by the individual library. In other words, the programs deal only with subfield a of the 050 field of the MARC format for books. The coder should include only classification numbers in the coding.

Abbreviating the L.C. number.

The second number of an L.C. range may be abbreviated if it has the same prefix as the first number. The prefix can only be broken off at specific points which are indicated by slashes in the following example:

AB/123/.A23

Sample abbreviations:

AB 120-123	OK (broken at first slash)
AB 120.3-.23	OK (broken at second slash)
AB 120-B 123	wrong
AB 120-23	wrong

Spacing

Leave one space after the first set of letters in an L.C. number.

General considerations.

Use the most recent edition of each L.C. schedule as well as the two Gale cummulated revisions of the schedules. If possible, peruse the quarterly supplements 1975-1978 to detect new subjects relevant to your HEGIS numbers.

Retain superseded L.C. numbers in the coding, because most libraries will not re-classify materials which were classified in these numbers.

3. Description.

Provide a text description of each L.C. number or range. Descriptions may be taken from the classification schedules, L.C. Subject Headings or constructed by the coder. They should represent current terminology, e.g. L.C.'s "deaf and dumb" should be changed to "hearing impaired".

If the same word is repeated several times in a subject, position that word first.

HISTORY, GREAT BRITAIN
HISTORY, SPAIN
HISTORY, U.S.

This sequence can be abbreviated on the coding form as:

HISTORY, GREAT BRITAIN
SPAIN
U.S.

Appendix 4

Instructions for Proof-reading Range Records

Introduction

One of the goals of the project is to produce a translation between the Library of Congress (LC) classification and the Higher Education General Information Survey (HEGIS) taxonomy for higher education academic programs. This will enable library collections classified in LC to be related to instructional programs which are classified in HEGIS.

The proof-reader will be working with range records, containing the LC class numbers and ranges and their descriptions which have been identified as supporting instructional programs (as represented by the HEGIS codes). Each range record contains the following fields:

1. HEGIS Code (four digits) identifying the instructional program
2. LC class number or range (25 characters)
3. Description of the LC class number or range (45 characters)

Proof-reading Tasks

1. Check that the syntax of the LC class number or range is correct. The computer must be able to parse them into the single number or beginning and ending numbers intended.

A range is coded as a single LC class number: Z 6432.P7
or as two LC numbers separated by a hyphen:

HA 1-100 (or) HA 1-HA100
HA 1.A1-.A32 (or) HA 1.A1-HA1.A32

The ending number of the range may be abbreviated if it has the same prefix as the starting number, but the prefix can only be broken off at the points indicated by slashes:

ABCD/1234/.A1234.B1678.C3012 etc.

Using the sample range from above,

HA 1.A1-1.A32	okay (broken at first slash)
HA 1.A1-.A32	okay (broken at second slash)
HA 1.A1-A1.A32	wrong (do not break LC letter group
HA 1.A1- 32	wrong these will catch HA 1.A1 through HA 32, and HA 2 respectively
HA 1.A1-2	wrong

2. Check that the LC class number or range appears in the LC schedules, and that the range description is a reasonable interpretation of that in the LC schedules. Some descriptions have been altered to reflect current usage, so they may not be identical. If the class number or range is not found in the schedules, or if there is a discrepancy between the ranges and schedules, check through the Additions and Changes. A large shelf list may also be consulted for verification of an LC number.

3. Mark with a question mark in left margin any single LC number which you suspect should be a range of numbers. For instance, if a single number HA 100 is given, but one or more decimal numbers which follow such as HA 100.2, HA 100.A-.Z seem to fit the same description, this line should be marked with a question mark. NOTE SUGGESTED CORRECTIONS IN PENCIL.

4. Check that all subdivisions of an LC range such as HA 100-100 pertain to the attached description. Mark these with a question mark if errant numbers or decimals seem to be included. Specify the errant items in pencil. Consider the program indicated by the HEGIS code, as some descriptions may have different meanings in different programs.

5. Correct spelling errors in the range descriptions.

6. Check that the range description is logical for the instructional program specified by the HEGIS code. (Do this generally during the proof-reading above, as well as on the separate HEGIS listing to be printed after all corrections are made to the LC ordered printout). Note: one particular LC number or range may be included in several different HEGIS numbers, as each HEGIS program should be considered individually, and not in relation to others.

7. For any and all records which require the attention of a staff member, put a question mark in the left margin in pencil, and either the actual correction or a brief description of the problem in the right margin.

Common Errors

1. Prefix of the end value of the LC range broken off at the wrong place: HA 9847-49 (the end value is less than the starting number).

2. Missing decimal points: BV 430.J5-L4 (should be -.L4). As this range exists, it would catch any book between BV 430.J5 and L 4.

3. O's and 0's mixed. (Alpha O vs. numeric 0, zero)

4. Transposition of numbers: HEGIS code 2202 for 2002.

5. Typos and misread letters: HG for HQ

The programs allow for 25 characters in the LC range, and eliminate any numbers beyond the 25 character limit. If range exceeds 25 characters, attempt to abbreviate and restore range meaning.

Proof-reading Procedure

Two computer listings of range records are printed. The first lists range records in order by LC class number, the second by HEGIS code and then class number.

Step 1. Compare the first listing to the LC schedules and perform proof-reading tasks 1 through 7. During these steps, mark all changes to be made on the first listing. This will serve as the rough copy from which changes will be subsequently coded for keypunching and input to the computer.

Step 2. The research team coordinator will review the first listing, making final decisions on all changes.

Step 3. The proof-reader will code range changes as described in the following section.

Step 4. The proof-reader will turn coding changes over to the programmer-analyst who will have the changes key-punched and run against the existing file of range records.

Step 5. The proof-reader will perform proof-reading task (p.) comparing the HEGIS code to the range description and indicating with a question mark in the left margin those ranges which do not appear logical. Also make note of HEGIS categories which seem incomplete.

Step 6. The research team coordinator will review the suspect ranges from Step 5. and assign new HEGIS numbers or verify accuracy of current HEGIS number.

Step 7. Proof-reader will code the necessary changes as described in the following section and turn them over to the programmer-analyst for keypunching.

Coding Range Record Changes

Changes are coded on the standard coding form. Each line of the form represents a change record.

The changes may be either typed or block-lettered; cursive writing is not permitted. HEGIS codes must be exactly four digits long. LC class numbers and ranges must be coded in twenty-five (25) characters or less. Descriptions of LC class numbers or ranges must not exceed forty-five (45) characters in length.

The following are the most common changes the proof-reader will need to make:

1. Correct an LC number or range, or HEGIS code
2. Correct a description
3. Add a new range record
4. Delete an existing range record

1. To correct an LC class number or range, or HEGIS code

Code two lines. On the first line, code the HEGIS code and LC class number or range of the incorrect record, exactly as given, and put either a D or DELETE in the description field. On the second line, code the correct LC class number or

range, the correct HEGIS code, and the correct description. Example: given the incorrect record 0835 CV 516-17 SPORTS, the corrections would be:

0835GV 516-17 D
0835GV 516-517 SPORTS

On the standard coding form, the HEGIS code goes in col. 7-10, the LC class number or range in col. 11-35, and the description in col. 36-80.

2. To correct a description

Code one line with the HEGIS code and LC class number or range from the range record being corrected, exactly as written, changing only the incorrect description. Example: given 1510 BS 143-689 HISTORY OF THE BIDLE, the correction would be:

1510BS 143-689 HISTORY OF THE BIBLE

There is no need to delete the previous record because the new range supersedes the old automatically.

3. To add a new range record

Code a line containing the HEGIS code, LC class number or range, and description of the new range record.

4. To delete an existing range record

Code a line with the HEGIS code and LC class number or range from the range record being deleted, exactly as written, putting D or DELETE in the description field. Example: given 9979 DDT 99996-2 UTTER NONSENSE, the deletion would read:

9979 DDT 99996-2 D (or)
9979 DDT 99996-2 DELETE

APPENDIX 5

HEGIS CLASSES EXPRESSED AS LC CLASS

1700	MATHEMATICS	
1701	MATHEMATICS, GENERAL	
	QA 1-99	General Mathematics
	QA 101-141	Arithmetic
	QA 150-271	Algebra
	QA 273-273.67	Probabilities
	QA 274-275	Stochastic Processes
	QA 297-433	Analysis
	QA 440-699	Geometry
	QA 801-939	Analytic Mechanics
	Z 6651-6655	Bibliography, Mathematics
1702	STATISTICS	
	BC 131-135	Symbolic & Mathematical Logic
	BC 141	Logic of Chance, Probability
	QA 276-295	Mathematical Statistics
	Z 7551-7556	Bibliography, Statistics
1703	APPLIED MATHEMATICS	
	BF 39	Statistical Method in Psychology
	D 16.17	Statistical Method in History
	HA 29-33	Statistical Method in Social Sciences
	HD 1421-1425	Agricultural Statistics
	HD 4915-4916	Wages, Statistical Theory
	HF 5691-5716	Business Math
	LB 2846-2846	Education Statistics

.

.

.

FORMAT 2
REVERSE ORDER INDICATING
MULTIPLE ALLOCATION OF LC RANGES
TO DIFFERENT HEGIS CODES

QA 269-272		GAMES	
	0507	OPERATIONS RESEARCH	78
	0913	INDUSTRIAL MANAGEMT ENG	042078
QA 273		PROBABILITY	
	0913	INDUSTRIAL MANAGEMT ENG	080978
QA 273-273		LOGIC OF CHANCE. PROBABILITY	
	1509	PHILOSOPHY	032978
QA 273-273.67		PROBABILITY	
	0503	BUSINESS STATISTICS	78
	1701	MATHEMATICS, GENERAL	042178
QA 273-274		PROBABILITIES	
	2405	RESOURCES MANAGEMENT	071378
	2412	FOREST RESEARCH	071378
QA 273-275		PROBABILITY & STOCHASTIC PROCESSES	
	2007	STATISTICS IN PSYCHOLOGY	050378
QA 273-295		MATHEMATICAL STATISTICS & PROBABILITY	
	0202	ARCHITECTURE	042078
	0419	BIOMETRICS &BIOSTATISTICS	050378
	0804	LU STATISTICS & RESEARCH	050378
	0825	LU TESTING,EVAL,MEASURES	081078
	1505	LINGUISTICS	050378
	1601	PHYSICAL SCIENCES, GEN	050378
	1902	PHYSICS,GEN-NU BIOPHYSCS	050378
	1905	CHEMISTRY,GEN-NU BIOCHEM	050378
	1911	ASTRONOMY	050378
	1914	GEOLOGY	050378
	1917	EARTH SCIENCES,GEN	050378
	2102	PUBLIC ADMINISTRATION	050378
	2103	PARKS & RECREATION MGT	050378
	2104	SOCIAL WORK & HELPNG SER	050378
	2105	LAW ENFORCE & CORRECTION	050378
	2202	ANTHROPOLOGY	050378
	2203	ARCHEOLOGY	050378
	2204	ECONOMICS	050378
	2205	HISTORY	050378
	2206	GEOGRAPHY	050378
	2208	SOCIOLOGY	050378
	2209	CRIMINOLOGY	050378
	2210	URBAN STUDIES	050378
	2212	DEMOGRAPHY	050378
QA 274-274.76		STOCHASTIC PROCESSES	
	0503	BUSINESS STATISTICS	081078
QA 274-275		STOCHASTIC PROCESSES	
	0913	INDUSTRIAL MANAGEMT ENG	080978
	1701	MATHEMATICS, GENERAL	78

Formula for Current Library Acquisitions

Purpose: A formula which will ensure an adequate flow of currently published materials into the different types of libraries in the University. The formula does not, except for the three emerging colleges, address the problem of retrospective acquisitions.

RATIONALE & DEFINITIONS

1. Basic Principles Followed:

   a. Current acquisitions

   A formula which is primarily responsive to the different campus missions, and secondarily to the different campus sizes as reflected by F.T.E.

   The formula places emphasis on careful selection of library materials from the current publishing output by each campus. The formula also recognizes the need for some duplication of heavily used materials caused by large enrollments.

   b. Retrospective collections

   The allocations are designed to maintain the currency of the library collections. An additional allocation is made which will support retrospective collecting for the emerging campuses only, at this time.

2. Elements and Values Used:

a. Current Collection Development

1. Base - Each campus library needs a base volume allocation used (a) to build and maintain its reference collections and (b) to round out the collection in areas not specifically allocated to a NEGIS discipline but nonetheless essential to students and faculty.

The reference collection base is the most vital segment of the library collection, sice it provides quick guidance to disciplinary and bibliographic information not maintained in depth in the library collection. More important, through published union lists, bibliographies, etc., it provides quick location access for interlibrary loan purposes to material not housed in the library.

For the colleges, 4800 volumes, or approximately 10% of the annual published output of trade and technical press, is used.

For the Ag&Techs, 2500 volumes, or 5% of the output, is used.

The specialized colleges, which have a narrower subject emphasis, are in close proximity to other major library resources and have smaller enrollments, are allocated a base of 1000 volumes.

2. Increments on Base - Campuses with larger enrollments find students competing for heavily used items and thus they need to acquire multiple copies of some titles to satisfy the demand.

The allocation of 100 volumes per 1000 (or fraction) FTE above a base of 4000 for colleges and 2000 for Ag & Techs enables the library to meet this need.

3. Disciplinary Allocation - It is now becoming recognized that in college level collections 20% of the collections will satisfy some 80% of the demand made on the collection.

Careful selection procedures by the library will thus answer the majority of immediate needs, but the allocation recognizes that the library will need to go to other resources via interlibary loan for rarer or more specialized books.

The immediate needs of graduate students for a slightly larger variety of material are met with an expanded NYCIS allocation of 25% per discipline.

The Ag & Techs have a different mission and a different emphasis in teaching from the four-year colleges and the specialized colleges. Their concerns with hands-on vocational education, the emphasis on specialties, and the relatively lower liberal arts component in turn reduces the emphasis on books.

The library is still, however, a vital support to the campus mission. The differences are recognized by a slightly lower base of 2500 volumes per year, whilst maintaining the increments for large enrollments by allocating 100 volumes per FTE above 2000 FTE. Negis disciplinary collections are maintained at 2 1/2% of current publishing output.

4. Specialized Colleges - The specialized colleges have library needs which are different from the other types of campus. They are closer to major library resources and do not in general have the spread of Negis disciplines found in the other campuses. At the same time, they are meeting many of the same needs as other libraries for their students and faculty and need a core reference collection.

A lower core of 1000 volumes is established. For general supporting Negis disciplines an allocation of 20% per discipline is made. For the heavy disciplinary concentrations, in which the libraries are building research collections in specialized fields, which are a state-wide resource, the allocation per Negis disciplinary is up to 80% per discipline of current published output.

5. Medical Schools - (i.e. Upstate and Downstate) Formula based on 100% of medicine output plus 25% Biological sciences output.

6. Publishing Output - Based on average volume output of American Book Publishing by discipline over years 1972-1975, Trade and Technical section.

Plus 15% to allow for materials excluded from the published statistics (i.e. some government and institutional publications, and books published in Canada, U.K., Europe, etc., wich may be needed to support the College programs).

Differential unit costs per MEGIS discipline, reflecting different needs, (e.g., artbooks which are expensive to manufacture) have been used in computation of dollar allocation.

7. Serials & A/V Materials - Current collecting patterns as reflected in present budget requests are maintained and added to monographs & request.

b. Retrospective collection development - emerging campuses

For the three emerging colleges, an additional 20% of total monographs is made to enable the libraries to sustain their program of retrospective acquisition during the period of campus establishment.

3. Acquisitions Disciplinary - based Formulae

UNIVERSITY CENTRES

Base	40000 volumes based on current publishing output
Plus	additional volumes per discipline concentration, selected advanced graduate programs (2000 volumes per program)
Plus	additional titles for advanced graduate professional schools or subjects (varying numbers of volumes per discipline) (see attached)
Plus	addition for undergraduate students (1000 volumes per each 2000 students over 5000)
Plus	addition for sponsored research
Plus	access factor to other research libraries

(N.B. Serials subscriptions included in volume totals)

(See Voigt, M. Acquisition rates in University Libraries, Coll. Res. Lib., July 1975 p 263-271)

(Also, see attached)

COLLEGES - RETRENCHING

Base 4000 Volumes

Plus 100 Volumes per 1000 FTE above 4000 or fraction of 1000

Plus 20% Current Publishing output (in volumes per undergraduate NCGIS discipline represented on the campus)

or 25% Current Publishing output (in Volumes per Graduate NCGIS discipline)

= Total Volumes

X Unit Cost per Volume per NCGIS disciplines

= Total for monographs

+ Serials $ request 1977-8

+ A/V $ request 1977-8

= Total $ for acquisitions

COLLEGES - EXPANDING

Base 4000 Volumes

Plus 100 Volumes per 1000 FTE above 4000 or fraction of 1000

Plus 20% Current Publishing output (in volumes per undergraduate NCGIS discipline represented on the campus)

or 25% Current Publishing output (in Volumes per Graduate NCGIS discipline)

Plus 20% Total monographs allocated as result of above computation. (Retrospective acquisition)

= Total Volumes

X Unit Cost per Volume per NCGIS disciplines

= Total for monographs

+ Serials $ request 1977-8

+ A/V $ request 1977-8

= Total $ for acquisitions

AG & TECHS. (as above with following values)

Base 2500 Volumes

Plus 100 Volumes per 1000 FTE above 2000 or fraction of 1000

Plus 2 1/2% Current Publishing output in Volumes per SECIS discipline

Plus Serials and A/V

SPECIALIZED COLLEGES (Forestry, Maritime, Optometry, Ceramics, Ag. Mil, Vet.)

Base 1000 Volumes

Plus 2 1/2% current publishing output per SECIS discipline (general)

Plus Additional support for selected SECIS disciplinary concentration up to 8% current published output. (Research collection support).

Plus Serials and A/V